MTH 254 Lecture Notes

Zachary Winkeler

Spring 2023

Contents

Chapter 1

Counting Principles

1.1 Introduction

The field of combinatorics has many sub-fields. Here are a few of them:

- **Enumerative combinatorics**: The mathematics of counting "things", where "things" usually means arrangements of objects and mathematical patterns. The goal is to find an explicit formula that counts these objects, but less-than-explicit formulas are also nice.
- **Analytic combinatorics**: Like enumerative combinatorics, analytic combinatorics is interested in counting stuff, but they care more about asymptotic formulas, i.e. "How quickly does the count grow when $n \to \infty$?"
- **Algebraic combinatorics**: The mathematics of counting algebraic objects like symmetric functions and ordered sets. Also, the study of ways to use algebraic structures (e.g. group actions) to count things.
- **Infinitary combinatorics**: A mixture of combinatorics and set theory, infinitary combinatorics asks questions about structures like infinitely-large graphs and trees and large cardinals.
- **Probabilistic combinatorics**: Asks questions like "What is the average number of triangles in a random graph?" Uses the probabilistic method, a method of proving the existence of an object with certain properties using probability.

There are many others, e.g. arithmetic combinatorics, combinatorial design theory, extremal combinatorics, geometric combinatorics, graph theory, matroid theory, order theory, partition theory, topological combinatorics, etc.

Question 1.1.1. Pop quiz![1]

(a) How many ways are there to pick 1 student from 6 first-years and 8 sophomores?

(b) How many ways are there to pick 1 piece of fruit from 6 oranges and 8 apples?

(c) How many ways are there to pick 1 letter from 3 A's, 5 B's, and 7 C's?

(d) How many ways are there to pick 2 letters from 3 B's and 3 G's?

(e) How many ways are there to pick 2 students from 3 juniors and 3 seniors?

(f) How many ways are there to pick 5 oranges from 6 oranges?

(g) How many ways are there to pick 5 professors from 6 professors?

(h) How many ways are there to pick 1 professor from 6 professors?

(i) How many ways are there to pick 5 pieces of fruit from 7 oranges and 8 apples?

(j) How many ways are there to pick some pieces of fruit from 8 oranges and 9 apples if at least one piece is picked?

[1] Adapted from [Mar01]

Let's talk about a few basic principles of counting.

- **The sum principle**: If A and B are disjoint sets ($A \cap B = \emptyset$), then

  $$|A \cup B| = |A| + |B| .$$

  We may use the notation $A \sqcup B$ instead to remind ourselves that A and B have no elements in common.

- **The product principle**: If A and B are sets, then

  $$|A \times B| = |A||B| .$$

- **The quotient principle**: If a set A is partitioned into n subsets, each of size m, then

  $$n = \frac{|A|}{m} .$$

  This is basically just a rephrasing of the product principle, but is useful for "fixing" answers where you have over-counted a set by a constant multiple.

- **The bijection principle**: If $f : A \to B$ is a bijection, then

  $$|A| = |B| .$$

Question 1.1.2. (a) How many ways are there to order a collection of n different objects?

(b) How many ways are there to make an ordered list of k objects from a collection of n different objects?

(c) How many ways are there to select k objects from a collection of n objects, if the order of selection is irrelevant?

Here are some symbols we will use to abbreviate common operations:

- The **factorial**
$$n! = n(n-1)(n-2)\dots(1)\,.$$
- The **falling factorial**
$$n^{\underline{k}} = n(n-1)(n-2)\dots(n-k+1)\,.$$
- The **rising factorial**
$$n^{\overline{k}} = n(n+1)(n+2)\dots(n+k-1)\,.$$

Question 1.1.3. Simplify each of the following:

(a) $5!$

(b) $5^{\overline{3}}$

(c) $5^{\underline{3}}$

(d) $n^{\underline{n}}$

(e) $1!$

(f) $1^{\overline{n}}$

(g) $0!$

(h) $n^{\overline{0}}$

(i) $n^{\underline{0}}$

(j) $n^{\underline{n-1}}$

The bijection principle seems simple, but it is also the key ingredient in **combinatorial proof**. It allows us to prove that two numbers are equal without even determining what those numbers are!

Question 1.1.4. Let A be any (finite) set. Prove that there are an equal number of subsets of A with even cardinality as there are subsets of A with odd cardinality, by defining a bijection between even- and odd-sized subsets.

1.2 Binomial coefficients

The set $[n]$ is defined to be $\{1, 2, \ldots, n\}$. The **binomial coefficient**

$$\binom{n}{k} = \frac{n^{\underline{k}}}{k!} = \frac{n!}{k!(n-k)!}$$

counts the number of k-element subsets of $[n]$, or really any n-element set. Equivalently, it counts the number of ways to choose k things from n (distinguishable) things, which is why it is often pronounced "n choose k".

Theorem 1.2.1.

$$\binom{n}{k} = \binom{n}{n-k}$$

Question 1.2.2. Give a combinatorial proof of Theorem 1.2.1, i.e. describe a bijection between a set of objects counted by $\binom{n}{k}$ and $\binom{n}{n-k}$.

A **bit string** is a sequence of 0's and 1's. We may denote the set of all bit strings of length n as $\{0, 1\}^n$ or $\{0, 1\}^{[n]}$. The **weight** of a bit string is the number of 1's in it.

Question 1.2.3. (a) How many bit strings have length n?

(b) How many bit strings have length n and weight k?

(c) How many bit strings have length n, weight k, and start with a 0?

(d) How many bit strings have length n, weight k, and start with a 1?

Pascal's triangle is a nice visual way to represent binomial coefficients.

$$
\begin{array}{ccccccccccccccc}
 & & & & & & & 1 & & & & & & & \\
 & & & & & & 1 & & 1 & & & & & & \\
 & & & & & 1 & & 2 & & 1 & & & & & \\
 & & & & 1 & & 3 & & 3 & & 1 & & & & \\
 & & & 1 & & 4 & & 6 & & 4 & & 1 & & & \\
 & & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & & \\
 & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 & \\
1 & & 7 & & 21 & & 35 & & 35 & & 21 & & 7 & & 1
\end{array}
$$

To be precise, $\binom{n}{k}$ appears as the kth element of the nth row.

Question 1.2.4. (a) Turn your answers to 1(b), 1(c), and 1(d) from the previous question into an identity involving binomial coefficients.

(b) Interpret your identity visually using Pascal's triangle.

Some remarks about Pascal's triangle:

- If you color only the odd numbers in Pascal's triangle, you get an interesting pattern![2]
- There are some similarities between the identity in 2(b) and the way you might think about derivatives in calculus.

[2] If you understand why this is, please explain it to me.

Question 1.2.5. Prove the following identity combinatorially, i.e. prove that both sides count the same set of objects in two different ways:

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n\,.$$

Question 1.2.6. Prove the following identity combinatorially:

$$\sum_{k=0}^{n} \binom{k}{m} = \binom{n+1}{m+1}$$

1.3 Multiset and multinomial coefficients

A **multiset** is like a set in which the same element can occur multiple times, e.g. $\{1, 2, 3\}$ and $\{1, 2, 2, 3, 3, 3\}$ are distinct multisets. The order of elements, however, still does not matter.

Question 1.3.1. An equivalent definition of a multiset is as a function $f : S \to \mathbb{N}$, where S is an actual set and $\mathbb{N}$ is the natural numbers.

(a) Explain why this is true.

(b) Translate the multiset $\{A, B, C, A, B, A\}$ into a function.

(c) Could we have defined ordinary sets/subsets in a similar way? Why or why not?

Question 1.3.2. How many multisets of size k can you make using the elements of $[n]$?

Your answer to question 2 is sometimes called the **multiset coefficient**, denoted:

$$\left(\!\!\binom{n}{k}\!\!\right).$$

Question 1.3.3. In how many ways can you put n distinguishable objects into m distinguishable boxes, such that k_1 objects end up in the first box, k_2 objects end up in the second box, and so on until k_m objects end up in the last box, where $n = k_1 + k_2 + \cdots + k_m$?

Your answer to question 3 is called a **multinomial coefficient**, denoted

$$\binom{n}{k_1, k_2, \ldots, k_m}.$$

Question 1.3.4. Generalize the following facts about binomial coefficients to multinomial coefficients.

(a) $\binom{n}{k} = \binom{n}{n-k}$

(b) $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$

(c) $\sum_{k=0}^{n} \binom{n}{k} = 2^n$

Question 1.3.5. How many anagrams of the word MISSISSIPPI are there? They don't need to be valid English words.

Question 1.3.6. (Open-ended) Is there a relationship between multisets and multinomial coefficients?

1.4 Problem session: Catalan numbers[3]

Question 1.4.1. In a part of a city, all streets run either north-south or east-west, and there are no dead ends. Suppose we are standing on a street corner. In how many ways may we walk to a corner that is four blocks north and six blocks east, using as few blocks as possible?

Question 1.4.2. The previous problem has a geometric interpretation in a coordinate plane. A **lattice path** in the plane is a "curve" made up of line segments that either go from a point (i, j) to the point $(i + 1, j)$ or from a point (i, j) to the point $(i, j + 1)$, where i and j are integers. (Thus lattice paths always move either up or to the right.) The length of the path is the number of such line segments.

(a) What is the length of a lattice path from $(0, 0)$ to (m, n)?

(b) How many such lattice paths of that length are there?

(c) How many lattice paths are there from (i, j) to (m, n), assuming i, j, m, and n are integers?

[3] This sequence of questions is adapted from [Bog+17].

Question 1.4.3. Notice that a lattice path from $(0,0)$ to (n,n) stays inside (or on the edges of) the square whose sides are the x-axis, the y-axis, the line $x = n$ and the line $y = n$. In this problem we will compute the number of lattice paths from $(0,0)$ to (n,n) that stay inside (or on the edges of) the triangle whose sides are the x-axis, the line $x = n$ and the line $y = x$. For example, in the figure below we show the grid of points with integer coordinates for the triangle whose sides are the x-axis, the line $x = 4$ and the line $y = x$.

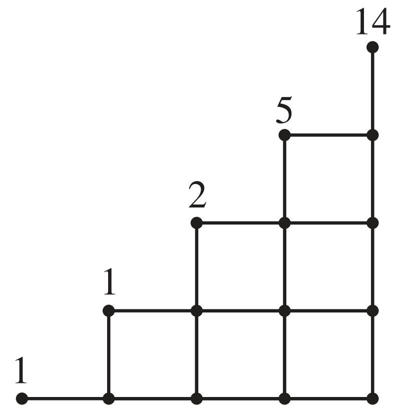


Figure 1.4.1: The lattice paths from $(0,0)$ to (i,i) for $i = 0,1,2,3,4$. The number of paths to the point (i,i) is shown just above that point.

(a) Explain why the number of lattice paths from $(0,0)$ to (n,n) that go outside the triangle is the number of lattice paths from $(0,0)$ to (n,n) that either touch or cross the line $y = x + 1$.

(b) Find a bijection between lattice paths from $(0,0)$ to (n,n) that touch (or cross) the line $y = x + 1$ and lattice paths from $(-1,1)$ to (n,n).

(c) Find a formula for the number of lattice paths from $(0,0)$ to (n,n) that do not go above the line $y = x$. The number of such paths is called a **Catalan number** and is usually denoted by C_n.

Your formula for the Catalan number can be expressed as a binomial coefficient divided by an integer. Whenever we have a formula that calls for division by an integer, an ideal combinatorial explanation of the formula is one that uses the quotient principle. Therefore, let's try to find one!

Question 1.4.4. Another kind of geometric path in the plane is a **diagonal lattice path**. Such a path is a path made up of line segments that go from a point (i, j) to $(i + 1, j + 1)$ (this is often called an **upstep**) or $(i + 1, j - 1)$ (this is often called a **downstep**), again where i and j are integers. (Thus diagonal lattice paths always move towards the right but may move up or down.)

(a) Describe which points are connected to $(0, 0)$ by diagonal lattice paths.

(b) What is the length of a diagonal lattice path from $(0, 0)$ to (m, n)?

(c) Assuming that (m, n) is such a point, how many diagonal lattice paths are there from $(0, 0)$ to (m, n)?

Question 1.4.5. A diagonal lattice path that never goes below the y-coordinate of its first point is called a **Dyck path**. We will call a Dyck path from $(0, 0)$ to $(2n, 0)$ a **Catalan path** of length $2n$. Thus the number of Catalan paths of length $2n$ is the Catalan number C_n.

(a) If a Dyck path has n steps (each an upstep or downstep), why do the first k steps form a Dyck path for each non-negative $k \leq n$?

(b) Thought of as a curve in the plane, a diagonal lattice path can have many local maxima and minima, and can have several absolute maxima and minima, that is, several highest points and several lowest points. What is the y-coordinate of an absolute minimum point of a Dyck path starting at $(0, 0)$? Explain why a Dyck path whose rightmost absolute minimum point is its last point is a Catalan path.

(c) Let D be the set of all diagonal lattice paths from $(0, 0)$ to $(2n, 0)$. (Thus these paths can go below the x-axis.) Suppose we partition D by letting B_i be the set of lattice paths in D that have i upsteps (perhaps mixed with some downsteps) following the last absolute minimum. How many blocks does this partition have? Give a succinct description of the block B_0.

(d) How many upsteps are in a Catalan path?

(e) We are going to give a bijection between the set of Catalan paths and the block B_i for each i between 1 and n. For now, suppose the value of i, while unknown, is fixed. We take a Catalan path and break it into three pieces. The piece F (for "front") consists of all steps before the ith upstep in the Catalan path. The piece U (for "up") consists of the ith upstep. The piece B (for "back") is the portion of the path that follows the ith upstep. Thus we can think of the path as FUB. Show that the function that takes FUB to BUF is a bijection from the set of Catalan paths onto the block B_i of the partition. (Notice that BUF can go below the x-axis.)

(f) Explain how you have just given another proof of the formula for the Catalan Numbers.

1.5 The pigeonhole principle

Theorem 1.5.1 (The Pigeonhole Principle)**.** *Let n and k be positive integers, and let $n > k$. Suppose we have to place n identical balls into k identical boxes. Then there will be at least one box in which we place at least two balls.*

The **pigeonhole principle** is perhaps the most obvious and intuitive counting principle; it is often used to come up with funny concrete examples like these:

- If 400 students take calculus this semester, then at least two calculus students share the same birthday.
- There are two people in the city of Phoenix, AZ that have the same number of hairs on their head.
- In a round-robin tournament, if everyone wins at least one match and loses at least one match, then at least two people have the same number of wins and losses.

Question 1.5.2. In each example above, which objects correspond to the "n balls" and which objects correspond to the "k boxes"?

However, it can be used to prove some unexpected *mathematical* results too!

Question 1.5.3. Prove that some element of the sequence $7, 77, 777, 7777, \ldots$ is divisible by 2023.

Theorem 1.5.4 (The Generalized Pigeonhole Principle)**.** *Let n and k be positive integers, and let $n > k$. Suppose we have to place n identical balls into k identical boxes. Then there will be at least one box in which we place at least $\lceil \frac{n}{k} \rceil$ balls.*

Question 1.5.5. Assume that there are 1.5 million residents of Phoenix, AZ and that no person has more than 250,000 hairs on their head. How does the generalized pigeonhole principle allow us to strengthen the result from the previous page?

Ten points are randomly chosen within a unit square.

(a) Prove that two of the points are closer than 0.48 units apart.

(b) Prove that three of the points can be covered by a disk of radius 0.5.

Here is another non-trivial application of the pigeonhole principle.

Theorem 1.5.6 (Dirichlet's Approximation Theorem)**.** *Suppose α is an irrational real number, and n is a positive integer. Then there exists a rational number $\frac{p}{q}$ with $1 \leq q \leq n$ satisfying*

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{q(n+1)}$$

Proof. Let $\{x\} = x - \lfloor x \rfloor$ denote the **fractional part** of a real number x. Consider the $n+2$ real numbers

$$0, \{\alpha\}, \{2\alpha\}, \ldots, \{n\alpha\}, 1\,.$$

Divide the interval $[0, 1]$ into $n + 1$ parts of equal length:

$$\left[0, \frac{1}{n+1}\right), \left[\frac{1}{n+1}, \frac{2}{n+1}\right), \ldots, \left[\frac{n}{n+1}, 1\right]\,.$$

By the pigeonhole principle, two of these fractional parts must be less than $\frac{1}{n+1}$ units apart from each other. Call them $\{k_1\alpha\}$ and $\{k_2\alpha\}$. Then:

$$\begin{aligned} \left|\{k_1\alpha\} - \{k_2\alpha\}\right| &< \frac{1}{n+1} \\ \big|\,(k_1\alpha - \lfloor k_1\alpha \rfloor) - (k_2\alpha - \lfloor k_2\alpha \rfloor)\,\big| &< \frac{1}{n+1} \\ \big|(k_1\alpha - k_2\alpha) - (\lfloor k_1\alpha \rfloor - \lfloor k_2\alpha \rfloor)\,\big| &< \frac{1}{n+1}\,. \end{aligned}$$

Note that $k_1 - k_2$ and $\lfloor k_1\alpha \rfloor - \lfloor k_2\alpha \rfloor$ are both integers; call the former q and the latter p. Then:

$$\begin{aligned} \big|q\alpha - p\big| &< \frac{1}{n+1} \\ \left|\alpha - \frac{p}{q}\right| &< \frac{1}{q(n+1)}\,. \end{aligned}$$

□

Question 1.5.7. (Challenge) A soccer team played 30 games this year, and scored a total of 53 goals, scoring at least one goal in each game. Prove that there was a sequence of consecutive games in which the team scored exactly six goals.

1.6 The inclusion-exclusion principle

Theorem 1.6.1 (Principle of Inclusion-Exclusion for Two Sets)**.** *Let A_1 and A_2 be finite sets. Then*

$$|A_1 \cup A_2| = |A_1| + |A_2| - |A_1 \cap A_2|$$

Question 1.6.2. How many numbers between 1 and 1000 are divisible by 3 or 5?

Question 1.6.3. Let A_1, A_2, and A_3 be finite sets.

(a) Write down a generalization of the inclusion-exclusion principle that counts $|A_1 \cup A_2 \cup A_3|$.

(b) Prove your answer to part (a).

Question 1.6.4. (Extra) If U and V are subspaces of a vector space, then

$$\dim(U+V) = \dim(U) + \dim(V) - \dim(U \cap V)\,.$$

Does your answer to 2(a) generalize to count $\dim(U+V+W)$ for subspaces U, V, and W?

Theorem 1.6.5 (Principle of Inclusion-Exclusion)**.** *Let $A_1, A_2, \ldots, A_n$ be finite sets. Then*

$$|A_1 \cup A_2 \cup \cdots \cup A_n| = \sum_i |A_i| - \sum_{i<j} |A_i \cap A_j| + \cdots \pm |A_1 \cap A_2 \cap \cdots \cap A_n|$$

The right side of the above equation counts the alternating sum of k-fold intersections of sets. We could write it more concisely as

$$\sum_{k=1}^{n} (-1)^{k+1} \sum_{i_1<i_2<\cdots<i_k} |A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}|$$

or

$$\sum_{I \subseteq [n]} (-1)^{|I|+1} |\bigcap_{i \in I} A_i|\,.$$

Question 1.6.6. How many prime numbers are less than 120? (Don't just list all of them!)

Question 1.6.7. (a) How many different functions $f : [n] \to [n]$ are there?

(b) How many of the functions in part (a) are bijections?

A **permutation** of $[n]$ is a bijection $\sigma : [n] \to [n]$. An element $i \in [n]$ such that $\sigma(i) = i$ is called a **fixed point** of σ. A permutation without any fixed points is called a **derangement**. In other words, a derangement is a way to rearrange n objects such that no object ends up in the same place it started.

Question 1.6.8. List all permutations of [3], and identify which points are fixed by each.

Question 1.6.9. (a) How many permutations of [4] have the element 1 as a fixed point?

(b) How many permutations of [4] have the elements 1 and 2 as fixed points?

(c) How many permutations of [4] are derangements?

We can count the number of derangements of n by first counting the number of permutations of $[n]$ with at least k fixed points:

$$(n-k)!\binom{n}{k}.$$

Then, the inclusion-exclusion principle tells us how many permutations have *any* number of fixed points:

$$\begin{aligned}
\sum_{k=1}^{n}(-1)^{k+1}(n-k)!\binom{n}{k} &= \sum_{k=1}^{n}(-1)^{k+1}(n-k)!\frac{n!}{k!(n-k)!} \\
&= \sum_{k=1}^{n}(-1)^{k+1}\frac{n!}{k!} \\
&= n!\sum_{k=1}^{n}\frac{(-1)^{k+1}}{k!}
\end{aligned}$$

Finally, we can subtract the number of permutations with fixed points from the total number of permutations to get the number of derangements:

$$\begin{aligned}
n! - n!\sum_{k=1}^{n}\frac{(-1)^{k+1}}{k!} &= n! + n!\sum_{k=1}^{n}\frac{(-1)^{k}}{k!} \\
&= n!\left(1+\sum_{k=1}^{n}\frac{(-1)^{k}}{k!}\right) \\
&= n!\sum_{k=0}^{n}\frac{(-1)^{k}}{k!}
\end{aligned}$$

Note that we are sneakily changing the lower bound $k=1$ to $k=0$ in the last step so that we can absorb the $1+\dots$ in front of the summation!

Question 1.6.10. What is the approximate probability that a randomly-chosen permutation of $[n]$ is a derangement? Assume $n \gg 0$, i.e. n is large enough that it doesn't affect the probability very much.

Chapter 2

Generating Functions

2.1 Recurrence relations

Recall that the **Catalan number** C_n counts *Catalan paths*, i.e. lattice paths from $(0,0)$ to (n,n) that do not cross the line $y = x$. One method of calculating C_n is to count all paths from $(0,0)$ to (n,n), and then use a clever bijective argument to count paths that do cross $y = x$. Subtracting these two numbers results in the explicit formula

$$C_n = \frac{1}{n+1}\binom{2n}{n}.$$

While combinatorial proofs can be fun, they often require an annoying amount of creativity. The theory of *generating functions* will allow us to compute formulas like the above expression in a more straightforward way.

For example, we could try to break down a Catalan path into smaller paths. If we have a Catalan path P from $(0,0)$ to (n,n) that touches the line $y = x$ at some point (k,k), then $P = P_1P_2$ is the concatenation of two Catalan paths P_1 and P_2. If (k,k) is the *first* such time that the path returns to the line $y = x$, then P_1 looks like $P_1 = RP_1'U$, where P_1' is also a Catalan path! Therefore, we can partition the set of all Catalan paths by the first time the path returns to the line $y = x$ to obtain the equation

$$C_n = \sum_{k=1}^{n} C_{k-1}C_{n-k}.$$

The above is an example of a **recurrence relation**, which is an equation that describes how terms in a sequence depend on previous terms. A recurrence relation may have many **solutions**, which are sequences that satisfy the relation. Solutions differ by their **initial conditions**, which are fixed initial values of the sequence. For the Catalan numbers, the above equation and the initial condition $C_0 = 1$ are enough to uniquely determine the sequence C_n.

Here is a preview of the kind of proof that generating functions will allow us to write.

- We start by defining a power series $C(x)$ that will have the Catalan numbers as coefficients, i.e. $C(x) = \sum_{n\geq 0} C_n x^n$.
- We use the recurrence relation $C_n = \sum_{k=1}^{n} C_{k-1}C_{n-k}$ and initial condition $C_0 = 1$ to determine that $C(x)$ satisfies the equation $C(x) = 1 + xC(x)^2$.
- We use the quadratic formula to solve this equation for $C(x)$, which means that $C(x) = \frac{1\pm\sqrt{1-4x}}{2x}$. We check the limit of $C(x)$ as $x \to 0$ to determine that we want the $\pm$ sign to be $-$ in this case.
- We use the generalized binomial theorem to write

  $$\sqrt{1-4x} = 1 - 2\sum_{n\geq 1}\frac{1}{n}\binom{2(n-1)}{n-1}x^n.$$

  This lets us simplify our above formula for $C(x)$ as

  $$C(x) = \sum_{n\geq 0}\frac{1}{n+1}\binom{2n}{n}x^n.$$

Let's go back to the basics and talk about some common recurrence relations. A sequence a_n that satisfies a recurrence relation of the form

$$a_n = a_{n-1} + d$$

is called **arithmetic**. All solutions to this relation have the closed form

$$a_n = a_0 + dn\,.$$

The sum of the first n terms of an arithmetic sequence is

$$\begin{aligned}\sum_{k<n} a_k &= n\frac{a_0 + a_{n-1}}{2} \\ &= n\frac{2a_0 + d(n-1)}{2}\,.\end{aligned}$$

Note that, unless $a_0 = d = 0$, the above formula does not converge as $n \to \infty$. Therefore, the sum of all the terms of an arithmetic sequence is either ∞ or $-\infty$.

A sequence b_n that satisfies a recurrence relation of the form

$$b_n = rb_{n-1}$$

is called **geometric**. All solutions to this relation have the closed form

$$b_n = r^n b_0\,.$$

The sum of the first n terms of a geometric sequence is

$$\begin{aligned}\sum_{k<n} b_k &= \frac{b_0 - b_n}{1-r} \\ &= b_0\frac{(1-r^n)}{1-r}\,.\end{aligned}$$

If $|r| < 1$, the above formula converges as $n \to \infty$, so we can also deduce that the sum of all the terms of a geometric series is

$$\begin{aligned}\sum_{k} b_k &= \lim_{n\to\infty} b_0\frac{(1-r^n)}{1-r} \\ &= \frac{b_0}{1-r}\,.\end{aligned}$$

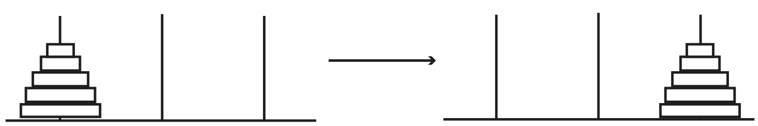
Question 2.1.1. (Challenge) Let c_n be a sequence satisfying a recurrence relation of the form $c_n = rc_{n-1} + d$. Find a closed-form expression for c_n in terms of c_0, n, r and d.

Usually, the most helpful tool for proving facts about recursively-defined things is *proof by induction*. For example, if we want to prove our formula for the sum of the first n terms of a geometric series, we could do so as follows:

- Base case: We verify that $\sum_{k<1} b_k = b_0 \frac{(1-r)}{1-r} = b_0$.
- Inductive step: We assume that $\sum_{k<n} b_k = b_0 \frac{(1-r^n)}{1-r}$ for some n, and then calculate that

$$\begin{aligned}
\sum_{k<n} b_k + b_n &= b_0 \frac{(1-r^n)}{1-r} + b_n \\
&= b_0 \frac{(1-r^n)}{1-r} + r^n b_0 \\
&= b_0 \frac{(1-r^n)}{1-r} + r^n b_0 \frac{1-r}{1-r} \\
&= b_0 \frac{1-r^n+r^n-r^{n+1}}{1-r} \\
\sum_{k<n+1} b_k &= b_0 \frac{1-r^{n+1}}{1-r}\,.
\end{aligned}$$

Question 2.1.2. The "Towers of Hanoi" puzzle has three rods rising from a rectangular base with n rings of different sizes stacked in decreasing order of size on one rod. A legal move consists of moving a ring from one rod to another so that it does not land on top of a smaller ring.[1]



(a) If H_n is the number of moves required to move all the rings from the initial rod to another rod that you choose, give a recurrence for H_n.

(b) Prove that $H_n = 2^n - 1$.

[1] Illustration from [Bog+17]

2.2 The generalized binomial theorem

Previously, we have defined the *binomial coefficient* $\binom{n}{k}$ as the number of ways to choose k objects from n identical objects. We also calculated closed forms for these numbers:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n^{\underline{k}}}{k!}\,.$$

Another (very related) way to interpret these numbers is as the coefficients of certain polynomials; this is known as the **binomial theorem**:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}\,.$$

We usually don't think of the variables x and y in the binomial theorem as actual numbers; instead, they are just symbols that we can add and multiply[2]. We might call expressions built using x and y **formal polynomials** to indicate that we are more interested in the polynomial itself than the function it represents.

Question 2.2.1. Prove the binomial theorem...

(a) ...combinatorially.

(b) ...by induction.

[2]In more algebraic terms, we could say that we are working over the ring $\mathbb{R}[x, y]$.

Question 2.2.2. (a) What is the coefficient of x^3 in $(x+1)^6$?

(b) What is the coefficient of x^3 in $x^2(x+1)^6$?

(c) What is the coefficient of x^2y^4 in $xy(x+2y)^4$?

Question 2.2.3. (a) What is the coefficient of x^2y^4z in $(x+y+z)^7$?

(b) Devise a generalization of the binomial theorem for expanding $(x+y+z)^n$.

One way to generalize the binomial theorem is to look at powers of polynomials with more than two terms. This naturally leads to the **multinomial theorem**, which states that

$$(x_1 + \cdots + x_m)^n = \sum_{k_1+\cdots+k_m=n} \binom{n}{k_1, \ldots, k_m} x_1^{k_1} \cdots x_m^{k_m}.$$

Another way to generalize the binomial theorem is to allow the power n to take on non-natural number values. While usual definition of the binomial coefficient

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

is only defined for $n, k \in \mathbb{N}$, the alternative definition as

$$\binom{n}{k} = \frac{n^{\underline{k}}}{k!}$$

makes sense for any $n \in \mathbb{R}$, with $k \in \mathbb{N}$ still required. Note that we will also require that $\binom{n}{k} = 0$ if $n, k \in \mathbb{N}$ and $k > n$. With these conventions in place, we can write down the **generalized binomial theorem**:

$$(x+1)^n = \sum_{k=0}^{\infty} \binom{n}{k} x^k$$

This theorem holds for any $n \in \mathbb{N}$ without restriction on x by the usual binomial theorem, and for any $n \in \mathbb{R}$ as long as $|x| < 1$.

One way of proving the generalized binomial theorem is via Taylor series. If $f(x) = (x+1)^n$, we compute its derivatives with respect to x to be

$$\begin{aligned}
f^{(0)}(x) &= (x+1)^n \\
f^{(1)}(x) &= n(x+1)^{n-1} \\
f^{(2)}(x) &= n(n-1)(x+1)^{n-2} \\
\vdots &= \vdots \\
f^{(k)}(x) &= n^{\underline{k}}(x+1)^{n-k}
\end{aligned}$$

and therefore can write

$$\begin{aligned}
(x+1)^n &= f(0) + xf'(0) + \frac{x^2}{2} f''(0) + \ldots \\
&= \sum_{k=0}^{\infty} \frac{x^k}{k!} f^{(k)}(0) \\
&= \sum_{k=0}^{\infty} \frac{n^{\underline{k}}}{k!} x^k
\end{aligned}$$

Note that some care is to be taken to distinguish between the case where $n \in \mathbb{N}$ and where $n \notin \mathbb{N}$, as in the former case, there are only finitely many nonzero derivatives to account for.

One surprising consequence of the generalized binomial theorem is that we obtain another proof of the formula for the sum of an infinite geometric sequence. If $n = -1$, then

$$\begin{aligned}
(x+1)^{-1} &= \sum_{k=0}^{\infty} \binom{-1}{k} x^k \\
\frac{1}{x+1} &= \sum_{k=0}^{\infty} \frac{(-1)^{\underline{k}}}{k!} x^k \\
&= \sum_{k=0}^{\infty} \frac{(-1)^k k!}{k!} x^k \\
&= \sum_{k=0}^{\infty} (-1)^k x^k
\end{aligned}$$

Substituting $x = -r$ into the above equation gives

$$\begin{aligned}
\frac{1}{1+(-r)} &= \sum_{k=0}^{\infty} (-1)^k (-r)^k \\
\frac{1}{1-r} &= \sum_{k=0}^{\infty} r^k
\end{aligned}$$

which becomes the usual formula after multiplying both sides by a_0.

Question 2.2.4. Find an algebraic proof that

$$\sum_{k \text{ even}} \binom{n}{k} = \sum_{k \text{ odd}} \binom{n}{k}.$$

Note that we already discussed a combinatorial proof of this fact the first week of class.

2.3 Formal power series and intro to generating functions

Formal power series

So far, we have discussed two ways of thinking about the equation

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \dots .$$

On one hand, it is a simple formula for calculating the sum of an infinite geometric series a_n with initial condition $a_0 = 1$ and constant ratio x (as long as $|x| < 1$). On the other hand, it is a special case of the generalized binomial theorem.

Another interpretation of the right-hand side is that it is a **formal power series** in x. A formal power series is any expression of the form

$$\sum_{i \geq 0} a_i x^i .$$

The set of all formal power series in x with real coefficients is denoted $\mathbb{R}[\![x]\!]$. The "formal" in formal power series indicates that we are not thinking of x as a real number and instead simply as a variable, and thus are not concerned with issues of whether or not the sum actually converges. With this in mind, we can interpret the left-hand side of the equation as telling us that the polynomial $1 - x$ is the *multiplicative inverse* of $1 + x + x^2 + \dots$. We can check this manually as well:

$$\begin{aligned}(1-x)(1+x+x^2+\dots) &= (1-x) + x(1-x) + x^2(1-x) + \dots \\ &= 1 - x + x - x^2 + x^2 - x^3 + \dots \\ &= 1\end{aligned}$$

One reason this is interesting is because not all formal power series (or even polynomials) have inverses! For example, x does not have an inverse as a formal power series. In algebra, we would say that $\mathbb{R}[\![x]\!]$ has the structure of a *ring*, but not a *field*, since we can add, subtract, and multiply any two elements together, but only divide them in certain situations.

Lemma 2.3.1. *A formal power series $\sum_{i \geq 0} a_i x^i$ has an inverse if and only if a_0 is non-zero.*

Question 2.3.2. Find the first 5 terms of the power series $\dfrac{x}{1 - x - x^2}$.

Intro to generating functions[3]

Suppose you are going to choose three pieces of fruit from among apples, pears and bananas for a snack. We can symbolically represent all your choices as

$$\text{🍎🍎🍎} + \text{🍐🍐🍐} + \text{🍌🍌🍌} + \text{🍎🍎🍐} + \text{🍎🍎🍌} + \text{🍎🍐🍐} + \text{🍐🍐🍌} + \text{🍎🍌🍌} + \text{🍐🍌🍌} + \text{🍎🍐🍌}$$

Here we are using a picture of a piece of fruit to stand for taking a piece of that fruit. Thus 🍎 stands for taking an apple, 🍎🍐 for taking an apple and a pear, and 🍎🍎 for taking two apples. You can think of the plus sign as standing for the "exclusive or," that is, 🍎 + 🍌 would stand for "I take an apple or a banana but not both." To say "I take both an apple and a banana," we would write 🍎🍌.

We can extend the analogy to mathematical notation by condensing our statement that we take three pieces of fruit to

$$\text{🍎}^3 + \text{🍐}^3 + \text{🍌}^3 + \text{🍎}^2\text{🍐} + \text{🍎}^2\text{🍌} + \text{🍎}\text{🍐}^2 + \text{🍐}^2\text{🍌} + \text{🍎}\text{🍌}^2 + \text{🍐}\text{🍌}^2 + \text{🍎🍐🍌}$$

In this notation 🍎^3 stands for taking a multiset of three apples, while $\text{🍎}^2\text{🍌}$ stands for taking a multiset of two apples and a banana, and so on. What our notation is really doing is giving us a convenient way to list all three element multisets chosen from the set $\{\text{🍎}, \text{🍐}, \text{🍌}\}$.

Suppose now that we plan to choose between one and three apples, between one and two pears, and between one and two bananas. In a somewhat clumsy way we could describe our fruit selections as

$$\begin{array}{lll} \text{🍎🍐🍌} + \text{🍎}^2\text{🍐🍌} + \dots & +\text{🍎}^2\text{🍐}^2\text{🍌} + \dots & +\text{🍎}^2\text{🍐}^2\text{🍌}^2 \\ \quad + \text{🍎}^3\text{🍐🍌} + \dots & +\text{🍎}^3\text{🍐}^2\text{🍌} + \dots & +\text{🍎}^3\text{🍐}^2\text{🍌}^2, \end{array}$$

but a much more succinct representation would be $(\text{🍎} + \text{🍎}^2 + \text{🍎}^3)(\text{🍐} + \text{🍐}^2)(\text{🍌} + \text{🍌}^2)$. Instead of using pictures of fruit, we could also use variables, e.g. A, P, and B.

Question 2.3.3. Substitute x for all the variables in the expression $(A + A^2 + A^3)(P + P^2)(B + B^2)$, and give an interpretation of the coefficient of x^n in terms of the fruit-related scenario discussed above.

[3]This sequence of questions is adapted from [Bog+17].

Question 2.3.4. (a) Write down a polynomial in the variable A that represents a choice of any number from zero to three apples.

(b) Write down a polynomial in three variables A, P, and B that represents a choice of zero to three apples, zero to three pears, and zero to three bananas.

(c) Write down a polynomial in one variable x such that the coefficient of x^n is the number of ways to choose n pieces of fruit from three apples, three pears, and three bananas.

(d) Suppose an apple costs 20 cents, a banana costs 25 cents, and a pear costs 30 cents. What should you substitute for A, P, and B in your answer to part (b) in order to get a polynomial in which the coefficient of x^n is the number of ways to choose a selection of fruit that costs n cents?

Question 2.3.5. Suppose that we are choosing a subset of $[n]$.

(a) Write down a polynomial in the variable x_i that represents the choice between choosing or not choosing i to be in our subset.

(b) Write down a polynomial in the variables $x_1, \ldots, x_n$ that represents the choice of choosing a subset of $[n]$.

(c) What should we substitute for x_i in your answer to part (b) to get a polynomial such that the coefficient of x^k is the number of ways to choose a k-element subset of n? What theorem did we just re-prove?

2.4 Generating functions[4]

A **generating function** is a way of representing a sequence as a formal power series. More specifically, given any sequence (a_n), we define its **ordinary generating function** (OGF) $A(x)$ to be the formal power series

$$A(x) = \sum_{n=0}^{\infty} a_n x^n \,.$$

Note that if (a_n) is a finite sequence, then $A(x)$ is actually just a polynomial. Here are some examples of sequences and their associated generating functions.

Sequence	Terms	Generating function
$a_n = 1$	$1, 1, 1, 1, \ldots$	$\dfrac{1}{1-x}$
$a_n = r^n$	$1, r, r^2, r^3, \ldots$	$\dfrac{1}{1-rx}$
$a_n = n$	$0, 1, 2, 3, \ldots$	$\dfrac{x}{(1-x)^2}$
$a_n = \dbinom{m}{n}$	$\dbinom{m}{0}, \dbinom{m}{1}, \ldots$	$(1+x)^m$
$a_n = \dbinom{n}{k}$	$\dbinom{0}{k}, \dbinom{1}{k}, \ldots$	$\dfrac{x^k}{(1-x)^{k+1}}$
$a_n = \dfrac{1+(-1)^n}{2}$	$1, 0, 1, 0, \ldots$	$\dfrac{1}{1-x^2}$
$a_n = \dfrac{1}{n!}$	$1, 1, \dfrac{1}{2}, \dfrac{1}{6}, \ldots$	e^x
$a_n = \dfrac{1}{n}, \quad (n > 0)$	$0, 1, \dfrac{1}{2}, \dfrac{1}{3}, \ldots$	$\ln \dfrac{1}{1-x}$

[4]The tables in this section were inspired by [SF14].

The correspondence between sequences and generating functions also applies to operations on these objects. For example, if a sequence (a_n) has generating function $A(x)$ and another sequence (b_n) has generating function $B(x)$, then the sequence $(a_n + b_n)$ has generating function $A(x) + B(x)$. Here is a list of a few common sequence operations and how they interact with generating functions.

Sequences		OGFs	
Operation	Formula	Operation	Formula
sum	$a_n + b_n$	**sum**	$A(x) + B(x)$
convolution	$\sum_k a_k b_{n-k}$	**product**	$A(x)B(x)$
right shift	a_{n-1}	–	$xA(x)$
left shift	a_{n+1}	–	$\frac{A(x) - a_0}{x}$
partial sum	$\sum_{k \le n} a_k$	–	$\frac{A(x)}{1-x}$
–	$(n+1)a_{n+1}$	**differentiation**	$A'(x)$
–	$\frac{a_{n-1}}{n}$	**integration**	$\int_0^x A(t)\,\mathrm{d}t$
–	$c^n a_n$	–	$A(cx)$
–	$a_n - a_{n-1}$	–	$(1-x)A(x)$

Question 2.4.1. Find a generating function for each sequence.

(a) The geometric sequence $2, 10, 50, 250, \ldots$.

(b) The arithmetic sequence $-1, 3, 7, 11, \ldots$.

(c) The sequence $a_n = 2^n + 3n + 6$.

Question 2.4.2. Find an expression for the nth term of the sequence represented by each generating function.

(a) $A(x) = \dfrac{1 + x + x^2}{1 + x}$.

(b) $B(x) = (1 - x)^2 \ln \dfrac{1}{1 - x}$.

Question 2.4.3. Let $m \in \mathbb{N}$, and let (a_n) be the sequence with with generating function $(1-x)^{-m}$. Find a simple expression for the n-th term of the sequence a_n; your answer should be a single binomial coefficient with positive inputs!

2.5 The product principle for OGFs

Let's write down a formula for the product of two formal power series $A(x) = \sum_n a_n x^n$ and $B(x) = \sum_n b_n x^n$:

$$A(x)B(x) = \sum_n \left(\sum_k a_k b_{n-k} \right) x^n$$

The sequence $c_n = \sum_k a_k b_{n-k}$ is called the **convolution** of the sequences a_n and b_n. Each term c_n is the sum of all pairs of terms a_k and b_{n-k} such that the sum of their indices is n.

Suppose that we have two sets S_1 and S_2. Let $v_1 : S_1 \to \mathbb{N}$ and $v_2 : S_2 \to \mathbb{N}$ be two functions, thought of as assigning a "value" to elements of the sets. We can define a natural function $v : S_1 \times S_2 \to \mathbb{N}$ on the product $S_1 \times S_2$ by letting $v(x_1, x_2) = v_1(x_1) + v_2(x_2)$.

Now, let $A_1(x) = \sum_k |v_1^{-1}(k)| x^k$ be the generating function where the coefficient of x^n is the number of elements in S_1 with value n, and define $A_2(x)$ likewise. The **product principle for ordinary generating functions** says that the coefficient of x^n in $A_1(x)A_2(x)$ is the number of ordered pairs $(x_1, x_2) \in S_1 \times S_2$ with value $v_1(x_1) + v_2(x_2) = n$.

Our first example of this principle will be an interpretation of the following fact:

Lemma 2.5.1.

$$(1-x)^{-n} = \sum_k \left(\!\!\binom{n}{k}\!\!\right) x^k$$

Proof. We will use the generalized binomial theorem to expand the left-hand side.

$$\begin{aligned}
(1-x)^{-n} &= \sum_k \binom{-n}{k} (-x)^k \\
&= \sum_k \frac{(-n)(-n-1)\dots(-n-(k-1))}{k!} (-1)^k x^k \\
&= \sum_k (-1)^k \frac{(-1)^k (n)(n+1)\dots(n+k-1)}{k!} x^k \\
&= \sum_k \frac{(n+k-1)\dots(n+1)(n)}{k!} x^k \\
&= \sum_k \binom{n+k-1}{k} x^k
\end{aligned}$$

□

How should be interpret this seemingly magical appearance of the multiset coefficients? We first think of $\dfrac{1}{1-x} = 1 + x + x^2 + \dots$ as representing a choice of any number of some object. Then the product

$$\left(\frac{1}{1-x}\right)^n = \frac{1}{1-x} \cdot \dots \cdot \frac{1}{1-x} = (1 + x + x^2 + \dots) \dots (1 + x + x^2 + \dots)$$

represents making a choice of any number of objects n times, where the number of ways to choose k total objects is grouped up in the coefficient of x^k.

Example 2.5.2. Suppose that a semester at some university consists of n days. At the beginning of each semester, the Dean of Engineering designs the term in the following way. She splits the term into two parts. The first k days of the term will form the theoretical part of the semester, and the second $n-k$ days will form the laboratory part (here $1 \leq k \leq n-2$). Then she chooses one holiday in the first part, and two holidays in the second part. In how many different ways can she design the term with these constraints?

Solution. Let P_n denote the number of ways the Dean can plan the semester. One way to approach the problem is to account for 1) all the ways to break the semester into two parts, 2) all the ways to choose 1 holiday in the first part, and 3) all the ways to choose 2 holidays in the second part. Putting these three pieces together gives us

$$P_n = \sum_{k=1}^{n-2} \binom{k}{1}\binom{n-k}{2}.$$

While this is completely correct, it is not as nice as it could be; the answer actually has a closed form without a summation sign! One indicator that we can use generating functions to simplify this expression is the fact that we are taking the sum over a variable k of a product of an expression involving a k and another involving $n-k$. This bears a resemblance to the formula for the product of generating functions.

More specifically, let $P(x)$ be the generating function for the sequence P_n, i.e.

$$P(x) = \sum_n \left(\sum_{k=1}^{n-2} \binom{k}{1}\binom{n-k}{2} \right) x^n$$

Consider the generating function

$$A(x) = \frac{x}{(1-x)^2}$$

for the sequence $a_n = \binom{n}{1}$, and the generating function

$$B(x) = \frac{x^2}{(1-x)^3}$$

for the sequence $b_n = \binom{n}{2}$. Then

$$\begin{aligned} P(x) &= A(x)B(x) \\ &= \frac{x}{(1-x)^2} \cdot \frac{x^2}{(1-x)^3} \\ &= \frac{x^3}{(1-x)^5} \\ &= x^3 \sum_n \binom{n+4}{4} x^n \\ &= \sum_n \binom{n+1}{4} x^n \end{aligned}$$

Therefore, $P_n = \binom{n+1}{4}$.

Question 2.5.3. Suppose a drawer contains three red beads, four blue beads, and five green beads. Use a generating function to determine the number of ways to select six beads if one must select at least one red bead, an odd number of blue beads, and an even number of green beads. *(Assume that beads of the same color are indistinguishable and that the order of selection is irrelevant.)*

Question 2.5.4. Suppose a drawer contains ten red beads, eight blue beads, and eleven green beads. Determine a generating function that encodes the answer to each of the following problems.

(a) The number of ways to select k beads from the drawer.

(b) The number of ways to select k beads if one must obtain an even number of red beads, an odd number of blue beads, and a prime number of green beads.

(c) The number of ways to select k beads if one must obtain exactly two red beads, at least five blue beads, and at most four green beads.

Question 2.5.5. Continuing from our semester planning example earlier, now assume that, instead of holidays, the Dean chooses some number of days for independent study in both parts of the semester.

(a) In how many different ways can she plan the semester with these constraints? Use generating functions to solve this problem.

(b) Interpret your answer to part (a) combinatorially, i.e. explain how you might have arrived at the same answer using counting principles we have already discussed.

2.6 Partial fractions

Example 2.6.1. Recall our earlier problem about a Dean planning a semester consisting of n days. This time, she is split the semester into three parts, choosing any number of holidays in the first part, and odd number of holidays in the second part, and an even number of holidays in the third part. How many different ways can she do this?

As before, let P_n be the number of ways to plan the semester, and let $P(x)$ be the generating function for this sequence. We will decompose $P(x)$ as a product of three functions, each representing one of the three parts of the semester.

- The first function $A(x)$ should represent choosing any number of holidays among n days. The sequence a_n represented by $A(x)$ is the geometric sequence $1, 2, 4, 8, \ldots$, so $A(x) = \dfrac{1}{1-2x}$.
- The second function $B(x)$ should represent choosing an odd number of holidays among n days. Given an n-element set, there are 2^{n-1} subsets of odd cardinality, so $b_n = 2^{n-1}$ if $n \geq 1$ and 0 otherwise. This gives us the shifted generating function $\dfrac{x}{1-2x}$.
- The third function $C(x)$ represents choosing an even number of holidays among n days. Using the same logic as for $B(x)$, we get that $c_n = 2^{n-1}$ if $n \geq 1$, and 1 if $n = 0$. Therefore, $C(x) = 1 + \dfrac{x}{1-2x} = \dfrac{1-x}{1-2x}$.

This lets us write

$$\begin{aligned} P(x) &= A(x)B(x)C(x) \\ &= \frac{1}{1-2x} \cdot \frac{x}{1-2x} \cdot \frac{1-x}{1-2x} \\ &= \frac{x(1-x)}{(1-2x)^3}. \end{aligned}$$

Question 2.6.2. Use the generating function above to find a simple expression for P_n.

While we could stop with our previous expression for $P(x)$, we can also use the method of **partial fractions** to simplify our generating function further.

$$\begin{aligned}
\frac{x-x^2}{(1-2x)^3} &= \frac{D}{1-2x} + \frac{E}{(1-2x)^2} + \frac{F}{(1-2x)^3} \\
x - x^2 &= D(1-2x)^2 + E(1-2x) + F \\
&= D(1-4x+4x^2) + E(1-2x) + F \\
&= D - 4Dx + 4Dx^2 + E - 2Ex + F
\end{aligned}$$

Comparing the coefficients of x^2 on both sides, we see that $-1 = 4D$, so $D = -\frac{1}{4}$. This lets us compare coefficients of x to get that $1 = 1 - 2E$, so $E = 0$. Finally, comparing the coefficients of x^0 gives us that $0 = -\frac{1}{4} + F$, so $F = \frac{1}{4}$. Therefore,

$$\frac{x-x^2}{(1-2x)^3} = -\frac{1}{4}\cdot\frac{1}{1-2x} + \frac{1}{4}\cdot\frac{1}{(1-2x)^3},$$

so we can write our generating function $P(x)$ as

$$\begin{aligned}
P(x) &= -\frac{1}{4}\cdot\frac{1}{1-2x} + \frac{1}{4}\cdot\frac{1}{(1-2x)^3} \\
&= -\frac{1}{4}\left(\sum_n 2^n x^n\right) + \frac{1}{4}\left(\sum_n \left(\!\binom{3}{n}\!\right) 2^n x^n\right).
\end{aligned}$$

We can extract the coefficient of x^n to find that

$$\begin{aligned}
P_n &= -\frac{1}{4}\cdot 2^n + \frac{1}{4}\cdot\left(\!\binom{3}{n}\!\right) 2^n \\
&= \left(\!\binom{3}{n}\!\right) 2^{n-2} - 2^{n-2} \\
&= 2^{n-2}\left(\left(\!\binom{3}{n}\!\right) - 1\right) \\
&= 2^{n-2}\left(\binom{n+2}{n} - 1\right) \\
&= 2^{n-2}\left(\binom{n+2}{2} - 1\right) \\
&= 2^{n-2}\left(\frac{(n+2)(n+1)}{2} - 1\right) \\
&= 2^{n-3}((n+2)(n+1) - 2) \\
&= 2^{n-3}(n^2 + 3n) \\
&= n(n+3)2^{n-3}.
\end{aligned}$$

In general, partial fractions applies anytime we have a quotient of polynomials

$$\frac{P(x)}{Q(x)}$$

with $\deg P(x) \leq \deg Q(x)$. We start by factoring the denominator $Q(x)$ into pairwise coprime polynomials $Q_1(x) \ldots Q_n(x)$, then set up and solve the equation

$$\frac{P(x)}{Q(x)} = \frac{C_1}{Q_1(x)} + \cdots + \frac{C_n}{Q_n(x)}.$$

If $Q(x)$ has any repeated factors, i.e. if $Q_i(x)^{k_i}$ is a factor of $Q(x)$, then our equation will instead contain a term for each distinct power of $Q_i(x)$, i.e.

$$\frac{P(x)}{Q(x)} = \frac{C_1}{Q_1(x)} + \cdots + \left(\frac{C_{i,1}}{Q_i(x)} + \frac{C_{i,2}}{Q_i(x)^2} + \cdots + \frac{C_{i,k_i}}{Q_i(x)^{k_i}}\right) + \cdots + \frac{C_n}{Q_n(x)}$$

Question 2.6.3. Find the coefficient of x^n in $\dfrac{x}{(1+x)^2(1-x)}$.

Question 2.6.4. Find the coefficient of x^n in $\dfrac{1}{(1+2x)(2-x)}$.

Question 2.6.5. The Dean can't stop planning new ridiculous schedules. This time, she wants to split the n days of the semester into two parts. The first part needs to contain an even number of days, while the second part can have any number of days. Additionally, there will be no holidays chosen in the first part, but any number of holidays may be selected in the second part. In how many ways can this semester be planned?

2.7 Recurrence relations, revisited

Recurrence relations

Consider the sequence $(a_n) = 5, 10, 20, 40, \ldots$ defined recursively by $a_0 = 5$ and $a_n = 2a_{n-1}$. We can recognize this as a geometric sequence, and we know from past experience that it has generating function

$$A(x) = \frac{a_0}{1 - rx} = \frac{5}{1 - 2x}\,.$$

Let's derive this formula a new way!

Previously, we have thought of a recurrence relation as a way to compute a_n using previous terms of the sequence. Another way to think of a recurrence relation is as an equation involving different *shifts* of a sequence. This is helpful because we can translate shifts of sequences into multiples of generating functions.

For example, the recurrence relation $a_n = 2a_{n-1}$ with initial condition $a_0 = 5$ can be translated into generating functions as

$$\begin{aligned}
A(x) &= 5 + 2xA(x) \\
A(x) - 2xA(x) &= 5 \\
(1 - 2x)A(x) &= 5 \\
A(x) &= \frac{5}{1 - 2x}\,.
\end{aligned}$$

Idea. In general, if we have a sequence (a_n) defined recursively, we can solve for the generating function $A(x)$ by replacing a_{n-k} by $x^k A(x)$ in the recurrence relation and adding terms to satisfy our initial conditions.

Here's another example. Let $a_n = a_{n-1} + n$ and $a_0 = 0$. Then

$$\begin{aligned}
A(x) &= xA(x) + \frac{x}{(1-x)^2} \\
A(x) - xA(x) &= \frac{x}{(1-x)^2} \\
(1 - x)A(x) &= \frac{x}{(1-x)^2} \\
A(x) &= \frac{x}{(1-x)^3} \\
A(x) &= x\left(\sum_n \left(\!\binom{3}{n}\!\right) x^n\right) \\
A(x) &= \sum_n \binom{n+2}{2} x^{n+1} \\
A(x) &= \sum_n \binom{n+1}{2} x^n
\end{aligned}$$

The Fibonacci numbers

How many ways are there to tile a $1 \times (n-1)$ grid with single tiles and dominoes? How many ways are there to climb $n-1$ stairs taking either one or two steps at a time? The answer to both of these questions is the **Fibonacci number** F_n.

The Fibonacci numbers are defined recursively by the initial conditions $F_0 = 0$ and $F_1 = 1$, and the recurrence relation $F_n = F_{n-1} + F_{n-2}$. We can translate this into an equation involving the generating function $F(x)$:

$$F(x) = x + xF(x) + x^2F(x)\,.$$

Solving this equation for $F(x)$ gives the closed form

$$\begin{aligned} F(x) &= x + xF(x) + x^2F(x) \\ F(x) - xF(x) - x^2F(x) &= x \\ (1-x-x^2)F(x) &= x \\ F(x) &= \frac{x}{1-x-x^2}\,. \end{aligned}$$

The denominator of this rational function factors as $1 - x - x^2 = (x+\phi)(x-\phi^{-1})$, where $\phi = \frac{1}{2}(1+\sqrt{5})$. Therefore, we can use partial fractions to decompose $F(x)$:

$$\begin{aligned} \frac{-x}{x^2+x-1} &= \frac{A}{x+\phi} + \frac{B}{x-\phi^{-1}} \\ -x &= A(x-\phi^{-1}) + B(x+\phi) \end{aligned}$$

We can solve the equations $A\phi^{-1} = B\phi$ and $-1 = A + B$ to get that $A = -\frac{1}{\sqrt{5}}\phi$ and $B = -\frac{1}{\sqrt{5}}\phi^{-1}$, so

$$\begin{aligned} F(x) &= \frac{1}{\sqrt{5}}\left(\frac{-\phi}{x+\phi} + \frac{-\phi^{-1}}{x-\phi^{-1}}\right) \\ \sqrt{5}\cdot F(x) &= -\frac{1}{1+\phi^{-1}x} + \frac{1}{1-\phi x} \\ &= -\left(\sum_n (-\phi^{-1}x)^n\right) + \left(\sum_n (\phi x)^n\right) \\ &= -\left(\sum_n (-\phi^{-1})^n x^n\right) + \left(\sum_n \phi^n x^n\right) \\ &= \sum_n \left(\phi^n - (-\phi)^{-n}x^n\right) \\ F(x) &= \sum_n \frac{\phi^n - (-\phi)^{-n}}{\sqrt{5}}x^n \end{aligned}$$

This gives us the closed formula for the nth Fibonacci number:

$$F_n = \frac{\phi^n - (-\phi)^{-n}}{\sqrt{5}}\,.$$

Let's look at another example of solving a counting problem with recurrence relations.

Example 2.7.1. How many different ways are there to choose some numbers from $[n]$ such that no two consecutive numbers are chosen?

We have previously solved this problem for a specific quantity k of numbers to choose, but now let's try to find a formula for the case where we consider all possible values of k.

Let b_n be the number of ways to choose a subset of $[n]$ subject to the restriction, and let $B(x)$ be the OGF of b_n. In order to find a recurrence relation for b_n, we will imagine that we are choosing such a subset from $[n]$, starting with the largest numbers first. We consider two cases:

- If we put n in our subset, then we cannot put $n-1$ in it, so there are b_{n-2} ways to choose the rest of our subset.
- If we don't put n in our subset, then we have no restriction on putting $n-1$ in it, so there are b_{n-1} ways to choose the rest of the subset.

Putting these two cases together, we get that $b_n = b_{n-1} + b_{n-2}$. Our initial conditions are that $b_0 = 1$ and $b_1 = 2$; note that we needed to specify the second initial condition since it does not follow from our recurrence relation.

We could then use this recurrence relation to get a formula for the generating function $B(x)$. Note that we subtract 1 from $B(x)$ in the expression $x(B(x)-1)$ so that we do not over-count our initial condition $b_1 = 2$.

$$\begin{aligned}
B(x) &= x(B(x)-1) + x^2B(x) + 1 + 2x \\
B(x) &= xB(x) + x^2B(x) + 1 + x \\
B(x) - xB(x) - x^2B(x) &= 1 + x \\
(1 - x - x^2)B(x) &= 1 + x \\
B(x) &= \frac{1+x}{1-x-x^2}\,.
\end{aligned}$$

However, we do not need to do much more work, as $B(x)$ has the same denominator as our generating function $A(x)$ from the previous example with Fibonacci numbers. In fact, we can write $B(x) = \frac{1}{x}A(x) + A(x)$. Therefore, we get that $b_n = F_{n+1} + F_n = F_{n+2}$.

One interesting consequence of this example, combined with problem 3a) from Homework #1, is that the Fibonacci numbers are related to binomial coefficients! If we add up all the ways to choose k numbers from $[n]$ with no two adjacent, we should get the same number as b_n. Putting it all together, we get that

$$\sum_{k=0}^{\lceil n/2 \rceil} \binom{n-k+1}{k} = F_{n+2}\,.$$

Since we're already at three pages, might as well add a fourth.

$$
\begin{array}{ccccccccccccccc}
 & & & & & & & 1 & & & & & & & \\
 & & & & & & 1 & & 1 & & & & & & \\
 & & & & & 1 & & 2 & & 1 & & & & & \\
 & & & & 1 & & 3 & & 3 & & 1 & & & & \\
 & & & 1 & & 4 & & 6 & & 4 & & 1 & & & \\
 & & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & & \\
 & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 & \\
1 & & 7 & & 21 & & 35 & & 35 & & 21 & & 7 & & 1
\end{array}
$$

Question 2.7.2. Interpret the formula from the end of the previous page in terms of Pascal's triangle. Where are the Fibonacci numbers?

2.8 Problem session: Poker chips

Note: These questions involve stacking poker chips in a weird way, but each scenario is separate, i.e. none of the restrictions carry over into later problems unless specifically noted.

Question 2.8.1. You have a bunch of poker chips in three distinct colors: white, red, and blue. You have a particular way of stacking the chips: starting with an empty stack, you repeatedly either add one white chip, two blue chips, or two red chips. Let A_n be the number of different stacks of n chips this process can result in.

(a) Find a recurrence relation satisfied by A_n.

(b) Use your recurrence relation from part (a) to find an equation for the generating function $A(x)$ for A_n.

(c) Find a closed-form expression for A_n.

Question 2.8.2. You have a bunch of poker chips in three distinct colors. You want to stack them so that each chip is touching at least one other chip of the same color. Let P_n be the number of ways to make a stack of n chips satisfying this condition.

(a) Find a recurrence relation satisfied by P_n.

(b) Use your recurrence relation from part (a) to find an equation for the generating function $P(x)$ for P_n.

(c) Find a closed-form expression for P_n.

Question 2.8.3. This time, you only want to use two different colors of poker chip. As before, you want to stack them so that each chip is touching at least one other chip of the same color. Let Q_n be the number of ways to make a stack of n chips satisfying this condition.

(a) Find a recurrence relation satisfied by Q_n.

(b) Use your recurrence relation from part (a) to find an equation for the generating function $Q(x)$ for Q_n.

(c) Relate Q_n to another sequence we have studied. *(Feel free to use a calculator to compute a few terms of $Q(x)$ or find a closed form for you.)*

Question 2.8.4. Generalize your generating functions from the previous problems to the case where there are k distinct colors of poker chip.

More explicitly, let $k \in \mathbb{N}$ be some fixed number of colors, and let R_n be the number of ways to stack n poker chips such that each chip is touching at least one other chip of the same color. Find the generating function $R(x)$ for R_n.

Question 2.8.5. Let Y_n be the number of different stacks of n chips you can make with only white and blue chips, such that each stack contains more white chips than blue chips. Find a closed-form expression for Y_n.

Question 2.8.6. Let Z_n be the number of different stacks of n chips you can make with only white and blue chips, such that there are no groups of 3 or more consecutive blue chips.

(a) Find a recurrence relation for Z_n.

(b) Use your recurrence relation from part (a) to find an equation for the generating function $Z(x)$ for Z_n.

2.9 The composition principle for OGFs[5]

Let's start by considering the following counting problem:

> All n soldiers of a military squadron stand in a line. The officer in charge splits the line at several places, forming smaller (nonempty) units. Then he names one person in each unit to be the commander of that unit. Let H_n be the number of ways he can do this. Find a closed formula for H_n.

First, let's modify the problem a bit and assume that the officer is splitting the line into k nonempty units. This is the ideal scenario for the OGF product principle; we can find the generating function for ways to choose a commander from a unit of size m, then raise that to the power of k. Since there are m ways to choose a single element from a set of size m, the relevant OGF is $B(x) = \dfrac{x}{(1-x)^2}$. Therefore, if we are given a fixed number k of units, then the answer would be the coefficient of x^n in

$$\left(\frac{x}{(1-x)^2}\right)^k = \frac{x^k}{(1-x)^{2k}}\,.$$

How can we use this information to solve the original problem? We could take the sum over all possible values of k:

$$\begin{aligned} H(x) &= \sum_k \left(\frac{x}{(1-x)^2}\right)^k \\ &= 1 + \left(\frac{x}{(1-x)^2}\right) + \left(\frac{x}{(1-x)^2}\right)^2 + \left(\frac{x}{(1-x)^2}\right)^3 + \dots\,. \end{aligned}$$

This would give us the correct generating function, but it does not seem very easy to work with. Instead, notice that we have seen this kind of sum before; if $y = \dfrac{x}{(1-x)^2}$, then this is simply the geometric series

$$\frac{1}{1-y} = \frac{1}{1-\frac{x}{(1-x)^2}} = \frac{1}{\frac{(1-x)^2-x}{(1-x)^2}} = \frac{(1-x)^2}{(1-x)^2-x} = 1 + \frac{x}{1-3x+x^2}\,.$$

We can then factor the denominator as $x^2-3x+1 = (x-\alpha)(x-\beta)$, where $\alpha = \frac{1}{2}(3+\sqrt{5})$ and $\beta = \frac{1}{2}(3-\sqrt{5})$. The partial fraction expansion is

$$1 + \frac{x}{1-3x+x^2} = 1 + \frac{1}{\sqrt{5}}\cdot\frac{x}{x-\alpha} + \frac{1}{\sqrt{5}}\cdot\frac{x}{x-\beta}\,.$$

We can use this to find a closed-form expression for H_n:

$$H_n = \begin{cases} 1 & n = 0 \\ \frac{1}{\sqrt{5}}(\alpha^n - \beta^n) & n \geq 1 \end{cases}\,.$$

[5] Material for this section adapted from [Bon16].

The appearance of the infinite geometric series $\frac{1}{1-x}$ in the example problem was not coincidental, as explained by the following theorem.

Theorem 2.9.1. *Let a_n be the number of ways to make a certain type of choice from an n-element set (with $a_0 = 0$), and let h_n be the number of ways to split the set $[n]$ into some number of disjoint non-empty intervals, then make that same type of choice on each of the pieces. Let $A(x)$ and $H(x)$ be the associated OGFs. Then*

$$H(x) = \frac{1}{1 - A(x)}.$$

In the example of the military squadron, the disjoint non-empty intervals are the divisions of the line of solders into individual units. The sequence a_n is $0, 1, 2, 3, \ldots$, representing the choice of a commander for each unit.

This theorem gives us an interpretation for the composition of the OGFs $\frac{1}{1-x}$ and $A(x)$. What would happen if we replaced $\frac{1}{1-x}$ by a different function? What do compositions represent in general?

Let's think about a slight variation on our problem:

> All n soldiers of a military squadron stand in a line. The officer in charge splits the line at several places, forming smaller (nonempty) units. Then he names one person in each unit to be the commander of that unit. Finally, he chooses some of the units to be on night duty. Let G_n be the number of ways he can do this. Find a closed formula for G_n.

As before, we can approach this problem by first considering the case that the officer splits the line into exactly k units, for some k. Also as before, we get that the generating function that represents ways to choose a commander in each unit is

$$\left(\frac{x}{(1-x)^2}\right)^k.$$

This time, however we also want to account for the number of ways to choose some units to be on night duty; since we are considering k as a constant, we can just multiply our OGF by 2^k:

$$2^k \left(\frac{x}{(1-x)^2}\right)^k.$$

Now, we do the same trick again and sum over all possible values of k, to get

$$\begin{aligned} G(x) &= \sum_k 2^k \left(\frac{x}{(1-x)^2}\right)^k \\ &= 1 + 2\left(\frac{x}{(1-x)^2}\right) + 2^2\left(\frac{x}{(1-x)^2}\right)^2 + 2^3\left(\frac{x}{(1-x)^2}\right)^3 + \ldots. \end{aligned}$$

Again, this is not particularly simple or easy to work with, but it *can* be written as a composition of two simpler functions! If $y = \dfrac{x}{(1-x)^2}$, then

$$G(x) = \frac{1}{1-2y} = \frac{1}{1-\frac{2x}{(1-x)^2}} = \frac{1}{\frac{(1-x)^2-2x}{(1-x)^2}} = \frac{(1-x)^2}{(1-x)^2-2x} = 1 + \frac{2x}{1-4x+x^2}\,.$$

Skipping the usual tedious algebra, we use partial fractions to get that $G(x)$ has the form

$$G(x) = 1 + \sum_{n=1}^{\infty} \frac{1}{\sqrt{3}} \left((2+\sqrt{3})^n - (2-\sqrt{3})^n\right) x^n$$

which gives us the closed form

$$G_n = \begin{cases} 1 & n = 0 \\ \frac{1}{\sqrt{3}}\left((2+\sqrt{3})^n - (2-\sqrt{3})^n\right) & n \geq 1 \end{cases}.$$

This example leads us to the following generalization of our previous theorem.

Theorem 2.9.2 (The Composition Principle for OGFs)**.** *Let a_n be the number of ways to make one type of choice from an n-element set (with $a_0 = 0$), and let b_n be the number of ways to make a second type of choice (with $b_0 = 1$). Let g_n be the number of ways to split the set $[n]$ into some number of disjoint non-empty intervals, make the first type of choice within each interval, and make the second type of choice among the set of intervals. Let $A(x)$, $B(x)$, and $G(x)$ be the associated OGFs. Then*

$$G(x) = B(A(x))\,.$$

In our second example, the sequence $a_n = 0, 1, 2, 3, \ldots$ represents the choice of a commander for each unit, and the sequence $b_n = 1, 2, 4, 8, \ldots$ represents the choice of a subset of units for night duty.

Question 2.9.3. How is our second theorem a generalization of our first? Where does $\dfrac{1}{1-x}$ come from?

Remark. The language in our composition principle is a bit vague; while we will not go any deeper into the subject right now, the theory of *combinatorial species* is a formalization of these concepts, and worth looking into if you're interested in learning more about generating functions and/or category theory.

Question 2.9.4. One definition of the nth Fibonacci number F_n is as the number of ways to tile a $1 \times (n-1)$ grid by tiles of size 1×1 and 1×2. We have previously found the OGF for F_n to be

$$F(x) = \frac{x}{1 - x - x^2}$$

using a recurrence relation. Prove that the equation for $F(x)$ above is correct using the composition principle.

Question 2.9.5. Question 2.8.1 reads:

> You have a bunch of poker chips in three distinct colors: white, red, and blue. You have a particular way of stacking the chips: starting with an empty stack, you repeatedly either add one white chip, two blue chips, or two red chips. Let A_n be the number of different stacks of n chips this process can result in.

Use the composition principle to find the generating function $A(x)$ for A_n without using a recurrence relation.

2.10 Multivariate OGFs

Binomial coefficients

While we have so far only studied generating functions involving one variable (usually x), we can also think about multivariate functions as encoding interesting sequences! For example, the binomial coefficients $\binom{n}{k}$ satisfy the equation

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}.$$

We can think of this as a *recurrence relation* involving two indices. In other words, let $b_{n,k} = \binom{n}{k}$. Then the above equation simply says that

$$b_{n,k} = b_{n-1,k} + b_{n-1,k-1}$$

for all $n \geq 0$ and $1 \leq k \leq n-1$. We can add the initial conditions $b_{n,0} = b_{n,n} = 1$ to uniquely specify the binomial coefficients. We could think of $(b_{n,k})$ as a 2-dimensional generalization of a sequence, or simply a function $b : \mathbb{N} \times \mathbb{N} \to \mathbb{R}$.

Given such a sequence $b_{n,k}$, we define the *ordinary generating function* to be the formal power series in two variables

$$B(x,y) = \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} b_{n,k} x^k y^n .$$

In our specific example of the binomial coefficients, we have that

$$B(x,y) = \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \binom{n}{k} x^k y^n .$$

Let's use our recurrence relation to find a simpler form for $B(x,y)$. As in the single-variable case, we replace $b_{n-i,k-j}$ with the term $x^j y^i B(x,y)$, add initial conditions, then simplify:

$$\begin{aligned} B(x,y) &= yB(x,y) + xyB(x,y) + 1 \\ B(x,y) - yB(x,y) - xyB(x,y) &= 1 \\ (1-y-xy)B(x,y) &= 1 \\ B(x,y) &= \frac{1}{1-y-xy} \end{aligned}$$

This equation seems simple, but contains a lot of information! For example, we can interpret $B(x,y)$ as a geometric series in y with ratio $(x+1)$ to obtain another perspective on the binomial theorem:

$$B(x,y) = \frac{1}{1-y-xy} = \frac{1}{1-(x+1)y} = 1 + (x+1)y + (x+1)^2y^2 + (x+1)^3y^3 + \ldots$$

Question 2.10.1. Treat $B(x, y)$ as a power series in just x (assume y is constant) and find a formula for the coefficient of x^k. Where have you seen this formula before?

Multiset coefficients

Let's find a multivariate generating function $M(x,y)$ for the multiset coefficients $\left(\!\binom{n}{k}\!\right)$. One way to do so would be to use a recurrence relation, but we will use a different approach. We already know the generating function for the multiset coefficient $\left(\!\binom{n}{k}\!\right)$ with n fixed:

$$(1-x)^{-n} = \sum_{k=0}^{\infty} \left(\!\binom{n}{k}\!\right) x^k .$$

Therefore, we know that

$$\begin{aligned} M(x,y) &= \sum_{n=0}^{\infty}\sum_{k=0}^{\infty} \left(\!\binom{n}{k}\!\right) x^k y^n \\ &= \sum_{n=0}^{\infty}(1-x)^{-n}y^n \\ &= \sum_{n=0}^{\infty}\left(\frac{y}{1-x}\right)^n \\ &= \frac{1}{1-\frac{y}{1-x}} \\ &= \frac{1-x}{1-x-y} . \end{aligned}$$

Using this multivariate generating function, we can find the generating function for the multiset coefficient $\left(\!\binom{n}{k}\!\right)$ with k fixed instead of n. To do this, we solve for the coefficient of x^k in $M(x,y)$:

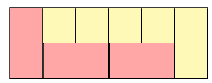
$$\begin{aligned} M(x,y) &= \frac{1-x}{1-x-y} \\ &= 1 + \frac{y}{1-x-y} \\ &= 1 + \frac{\frac{y}{1-y}}{1-\frac{x}{1-y}} \\ &= 1 + \frac{y}{1-y}\left(\sum_{k=0}^{\infty}\left(\frac{1}{1-y}\right)^k x^k\right) \\ &= 1 + \left(\sum_{k=0}^{\infty}\frac{y}{(1-y)^{k+1}}x^k\right) \\ &= \frac{1}{1-y} + \left(\sum_{k=1}^{\infty}\frac{y}{(1-y)^{k+1}}x^k\right) . \end{aligned}$$

Therefore, when k is a constant, the single-variable generating function (in y) for $\left(\!\binom{n}{k}\!\right)$ is $\frac{1}{1-y}$ if $k=0$, and $\frac{y}{(1-y)^{k+1}}$ if $k \geq 1$.

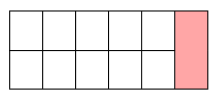
2.11 Multiple recurrence relations

A 2D Fibonacci Tiling

We know that the Fibonacci number F_n counts the ways to cover a $1 \times (n-1)$ grid with single tiles and dominoes. How many ways are there to cover a $2 \times (n-1)$ grid with the same kinds of tile? (Actually, for the sake of using sensible indices, let's just ask about covering a $2 \times n$ grid.)



Let a_n be the number of ways to tile a $2 \times n$ grid. Let's try to find a recurrence relation for a_n by splitting the tilings into cases based on what kind of tile covers the top right corner. It can either be a domino placed vertically, a single tile, or a domino placed horizontally. In the first case, we get a clear recursive relationship; there are a_{n-1} ways to tile the rest of the grid.

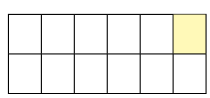


a_{n-1}

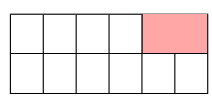
In the other two cases, though, it is not clear how to connect the ways to tile the rest of the grid back to tiling a $2 \times k$ rectangle for any value of k.

$$a_n = a_{n-1} + \text{ ???}$$

One solution to this problem is to create two more sequences to cover these cases. Let b_n be the number of ways to tile every square on a $2 \times n$ grid except the top-right corner, and let c_n be the number of ways to tile every square on a $2 \times n$ grid except the top-right corner and the square immediately to its left.



b_n



c_n

We can now find recurrence relations involving all three sequences:

$$\begin{aligned} a_n &= a_{n-1} + b_n + c_n \\ b_n &= a_{n-1} + b_{n-1} \\ c_n &= a_{n-2} + b_{n-1} \end{aligned}$$

and turn them into equations involving their generating functions:

$$\begin{aligned} A(x) &= xA(x) + B(x) + C(x) + 1 \\ B(x) &= xA(x) + xB(x) \\ C(x) &= x^2 A(x) + xB(x)\,. \end{aligned}$$

Since $A(x)$ is the generating function that we would like to find, we can solve for $B(x)$ in terms of $A(x)$ in the second equation, and $C(x)$ in terms of $A(x)$ in the third:

$$\begin{aligned}
B(x) &= xA(x) + xB(x) \\
B(x) - xB(x) &= xA(x) \\
(1-x)B(x) &= xA(x) \\
B(x) &= \frac{xA(x)}{1-x}
\end{aligned}$$

and

$$\begin{aligned}
C(x) &= x^2A(x) + xB(x) \\
C(x) &= x^2A(x) + x\left(\frac{xA(x)}{1-x}\right) \\
C(x) &= x^2A(x) + \frac{x^2A(x)}{1-x}\,.
\end{aligned}$$

Putting it all together, we get that

$$\begin{aligned}
A(x) &= xA(x) + B(x) + C(x) + 1 \\
A(x) &= xA(x) + \frac{xA(x)}{1-x} + x^2A(x) + \frac{x^2A(x)}{1-x} + 1\,,
\end{aligned}$$

which we can rearrange to get

$$\begin{aligned}
A(x) - xA(x) - \frac{xA(x)}{1-x} - x^2A(x) - \frac{x^2A(x)}{1-x} &= 1 \\
\left(1 - x - \frac{x}{1-x} - x^2 - \frac{x^2}{1-x}\right)A(x) &= 1 \\
\left(\frac{1 - x - x(1-x) - x - x^2(1-x) - x^2}{1-x}\right)A(x) &= 1 \\
\left(\frac{x^3 - x^2 - 3x + 1}{1-x}\right)A(x) &= 1\,,
\end{aligned}$$

so we are finally left with the simple closed-form expression

$$A(x) = \frac{1-x}{x^3 - x^2 - 3x + 1}\,.$$

Is there a closed-form expression for the coefficient a_n of x^n? There definitely is, since the denominator is a cubic polynomial and therefore has a solution in terms of radicals. However, it is not very nice, so we will leave this generating function as is. A computer will tell us that the first few terms of the sequence (a_n) are $1, 2, 7, 22, 71, 228, 733, \ldots$, and that the closed-form for a_n is

$$a_n = \frac{(\gamma-1)\gamma^{n+1}}{(\gamma-\beta)(\gamma-\alpha)} + \frac{(\alpha-1)\alpha^{n+1}}{(\alpha-\gamma)(\alpha-\beta)} + \frac{(\beta-1)(\beta^{n+1})}{\beta^2+2\alpha\gamma+1} \qquad \left(\begin{array}{l}\alpha,\ \beta,\text{ and }\gamma\text{ are the three}\\ \text{roots of } x^3-x^2-3x+1.\end{array}\right)\,.$$

Chapter 3

Distributions

3.1 Distribution problems

So far, many of our problems have involved counting ways to put objects in containers ("balls in boxes"), or at least have been equivalent to such a problem. This type of question is called a **distribution problem**.

One famous categorization of distribution problems is called the **twelvefold way**. This system characterizes "balls in boxes" problems based on the following three questions:

- Are the balls distinguishable?
- Are the boxes distinguishable?
- Are we required to put at most or at least one ball in each box?

The following table enumerates the ways to put n balls in k boxes in each scenario that we have studied so far.

Distinguishable?	No restriction	At most one ball per box	At least one ball per box
Balls and boxes	k^n	$k^{\underline{n}}$	
Only balls		$\begin{cases} 1 & n \le k \\ 0 & \text{otherwise} \end{cases}$	
Only boxes	$\left(\!\binom{k}{n}\!\right)$	$\binom{k}{n}$	$\left(\!\binom{k}{n-k}\!\right)$
Neither		$\begin{cases} 1 & n \le k \\ 0 & \text{otherwise} \end{cases}$	

Before we tackle the empty boxes in our table, let's take another look at the two multiset coefficients appearing in it. The left coefficient, $\left(\!\binom{k}{n}\!\right)$, counts ways to distribute identical balls among distinguishable boxes. It also counts the number of **weak compositions** of n with k parts, i.e. the number of ways to write n as a sum of k natural numbers. For example, if $n = 4$ and $k = 3$, then we have $\left(\!\binom{3}{4}\!\right) = \binom{6}{4} = 15$ ways to write 4 as a sum of 3 natural numbers.

$0+0+4$	$0+1+3$	$0+2+2$	$0+3+1$	$0+4+0$
$1+0+3$	$1+1+2$	$1+2+1$	$1+3+0$	$2+0+2$
$2+1+1$	$2+2+0$	$3+0+1$	$3+1+0$	$4+0+0$

You may find it helpful to think about the sums above as telling us how many balls go in each box.

If we ask how many ways there are to write n as a sum of k *non-zero* natural numbers, then we are just counting **compositions** of n with k parts. We can make a composition of n by first allotting 1 for each part, and then adding a weak composition; therefore, we have $\left(\!\binom{k}{n-k}\!\right)$ of these. If $n = 4$ and $k = 3$, then we have $\left(\!\binom{3}{4-3}\!\right) = \left(\!\binom{3}{1}\!\right) = \binom{3}{1} = 3$ of these; they are highlighted in the list above.

Question 3.1.1. (a) How many weak compositions of 4 with 6 parts are there?

(b) How many of the above weak compositions are also compositions?

Question 3.1.2. How many compositions of 7 with 4 parts are there, where the first part is not 3?

Question 3.1.3. How many compositions of n with k parts are there, where each part is ≥ 2?

Question 3.1.4. (a) How many compositions of n are there with any number of parts?

(b) How many weak compositions of n are there with any number of parts?

3.2 Partitions

Now, our goal is to tackle the remaining empty boxes of the last row of the twelvefold way: how to count ways to put n balls in k boxes, where neither type of object is distinguishable.

Recall that a *composition* of n with k parts is a way to write n as a sum of k non-zero natural numbers. With compositions, we care about the order of the summands. If we choose instead to forget the order of the summands, we get **partitions** of n with k parts. For example, there are 10 compositions of 6 with 3 parts:

$$\begin{array}{ccccc} 1+1+4 & 1+2+3 & 1+3+2 & 1+4+1 & 2+1+3 \\ 2+2+2 & 2+3+1 & 3+1+2 & 3+2+1 & 4+1+1 \end{array}$$

but only 3 distinct partitions of 6 with 3 parts:

$$1+1+4 \qquad\qquad 1+2+3 \qquad\qquad 2+2+2$$

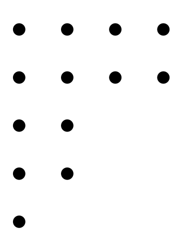
We use the notation $p_k(n)$ to denote the number of partitions of n with k parts. If we want to count all possible partitions of n into any number of parts, we would write simply $p(n)$; in symbols:

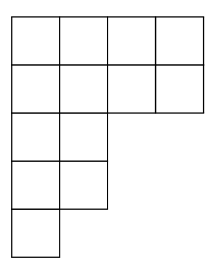
$$p(n) = \sum_k p_k(n)\,.$$

There is no nice closed-form expression for $p_k(n)$ or $p(n)$, however, we do have recurrence relations and generating functions. Despite their relative difficulty of computation, partitions are one of the most-studied distribution problems due to their rich structure and applications in other areas of math.

One way to view a partition is as a composition that is in (weakly) increasing order; we may write the partition as a k-tuple $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$. If λ is a partition of n, we denote this relationship by $\lambda \vdash n$.

Another way to visualize partitions is with diagrams. Here are two different types of diagrams representing the partition $4+4+2+2+1$; on the left is a **Ferrers diagram**, and on the right is a **Young diagram**:

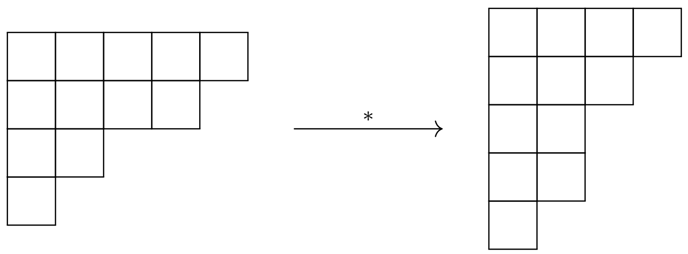




Note. These diagrams are read in such a way that each row is one of the parts of the partition; be aware that other sources may have different conventions as to how these are drawn!

These diagrams are more than just helpful visualization tools; they also reveal interesting *structure* in the space of partitions. For example, given any partition $\lambda \vdash n$, we can form the **conjugate** $\lambda^* \vdash n$ by reflecting the Young diagram of λ across the diagonal:

$$\lambda = (5, 4, 2, 1) \qquad\qquad \lambda^* = (4, 3, 2, 2, 1)$$



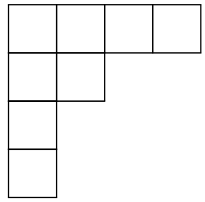
Conjugation can be a helpful tool for proving properties about partitions.

Lemma 3.2.1. *There are an equal number of partitions of n with k parts as there are partitions of n with greatest part equal to k.*

Proof. We can prove this by noting that conjugation gives us a bijection between these two sets. Any partition $\lambda \vdash n$ with k parts has k rows in its Young diagram, so $\lambda^* \vdash n$ has k *columns* in its Young diagram. The number of columns indicates the greatest part in λ^*, so the operation $\lambda \mapsto \lambda^*$ indeed gives a bijection between partitions of n with k parts and partitions of n with greatest part equal to k. $\square$

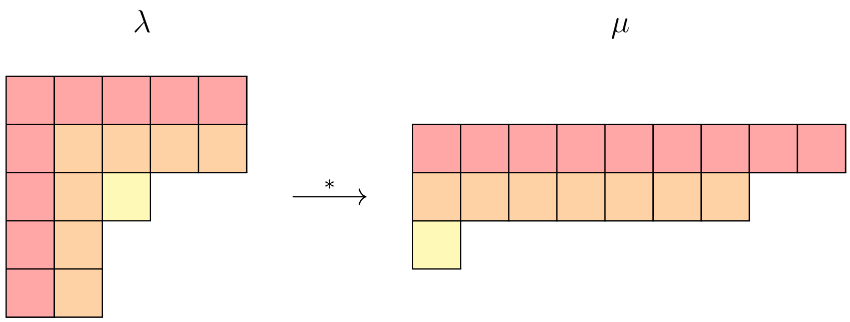
Corollary 3.2.2. *There are an equal number of partitions of n with $\leq k$ parts as there are partitions of n where every part is $\leq k$.*

Most of the time conjugation gives us a different partition from the one we started with, but some partitions are **self-conjugate**, like $(4, 2, 1, 1)$:



Lemma 3.2.3. *There are an equal number of self-conjugate partitions of n as there are partitions of n into distinct odd parts.*

Proof. We will again prove this lemma with a bijection, but this time it will be slightly more involved. Given a self-conjugate partition $\lambda \vdash n$, we want to break it up into "shells" as follows: the first shell consists of all boxes in the first row and/or first column, the second shell consists of all boxes in the second row and/or second column that haven't already been selected, and so on. We turn each shell of λ into a part of a new partition μ. For example, if $\lambda = (5, 5, 3, 2, 2)$, then $\mu = (9, 7, 1)$, as illustrated below.



Since λ is self-conjugate, each shell contains an odd number of boxes, and since Young diagrams are ordered in decreasing order top-to-bottom and left-to-right, each shell contains a *distinct* number of boxes. Therefore, μ is a partition of n into distinct odd parts. The correspondence $\lambda \mapsto \mu$ gives us a bijection between the set of self-conjugate partitions of n and the set of partitions of n into distinct odd parts, and therefore these two sets must have the same cardinality. $\square$

Generating functions

While there is no nice closed-form expression for $p(n)$, we could still ask for a closed-form for the *generating function* of $p(n)$. Unfortunately, this doesn't exist either!

Let $P(x)$ be the ordinary generating function of the sequence $p(n)$, i.e.

$$P(x) = \sum_n p(n)x^n .$$

Since partitions do not care about the order of their parts, we can use the product principle for OGFs to break down $P(x)$ as a product of generating functions for "partitions of n using only 1s", "partitions of n using only 2s", etc.

$$P(x) = (1 + x + x^2 + \dots)(1 + x^2 + x^4 + \dots)(1 + x^3 + x^6 + \dots)\dots$$

Each factor in the above expression can be simplified:

$$P(x) = \frac{1}{1-x} \cdot \frac{1}{1-x^2} \cdot \frac{1}{1-x^3} \cdot \dots$$

Ultimately, we get a (non-closed-form) expression:

$$P(x) = \prod_{i=1}^{\infty} \frac{1}{1-x^i} .$$

As far as non-closed-form expressions go, this one isn't bad, I guess. It could be a lot worse.

Question 3.2.4. (a) Find an expression for the generating function $P_{\leq k}(x)$ of the sequence $p_{\leq k}(n)$, the partitions of n into $\leq k$ parts for some fixed k.

(b) Find an expression for the generating function $P_k(x)$ of the sequence $p_k(n)$, the partitions of n into exactly k parts for some fixed k.

3.3 Problem session: Partitions

Since there is no closed-form expression for the number of partitions of n into k parts, we will simply fill in the relevant cell of the twelvefold way table with $p_k(n)$.

Ways to distribute n balls among k boxes

Distinguishable?	No restriction	At most one ball per box	At least one ball per box
Balls and boxes	k^n	$k^{\underline{n}}$	
Only balls		$\begin{cases} 1 & n \le k \\ 0 & \text{otherwise} \end{cases}$	
Only boxes	$\left(\!\binom{k}{n}\!\right)$	$\binom{k}{n}$	$\left(\!\binom{k}{n-k}\!\right)$
Neither		$\begin{cases} 1 & n \le k \\ 0 & \text{otherwise} \end{cases}$	$p_k(n)$

Question 3.3.1. (a) Prove that the number of partitions of n into at most k parts is equal to that of partitions of $n+k$ into exactly k parts.

(b) Fill in the bottom-left cell in the above table with an appropriate formula.

Question 3.3.2. Recall that a **lattice path** in the plane is a "curve" made up of line segments that either go from a point (i, j) to the point $(i + 1, j)$ or from a point (i, j) to the point $(i, j + 1)$, where i and j are integers.

(a) Draw all lattice paths from $(0, 0)$ to $(2, 2)$.

(b) Draw all Young diagrams with ≤ 2 rows and ≤ 2 columns.

(c) Let $L_{i,j}$ be the number of partitions (of any number) with $\leq j$ parts, each of which has size $\leq i$. Find a closed-form expression for $L_{i,j}$.

Question 3.3.3. (a) Draw the Young diagram for each partition of 8 into 4 parts.

(b) Draw the Young diagram for each partition of 24 into 4 parts, each of which has size ≤ 7.

(c) Show that the number of partitions of k into four parts equals the number of partitions of $3k$ into four parts of size at most $k-1$.

Question 3.3.4. Find a combinatorial proof (e.g. using Young diagrams) that the number of partitions of n into distinct parts is equal to the number of partitions of n with odd parts.

3.4 Set partitions

The last remaining cells of the twelvefold way involve distributions where the balls are distinct, and the boxes may or may not be.

Ways to distribute n balls among k boxes

Distinguishable?	No restriction	At most one ball per box	At least one ball per box
Balls and boxes	k^n	$k^{\underline{n}}$	
Only balls		$\begin{cases} 1 & n \leq k \\ 0 & \text{otherwise} \end{cases}$	
Only boxes	$\left(\!\binom{k}{n}\!\right)$	$\binom{k}{n}$	$\left(\!\binom{k}{n-k}\!\right)$
Neither	$p_k(n+k)$	$\begin{cases} 1 & n \leq k \\ 0 & \text{otherwise} \end{cases}$	$p_k(n)$

We will focus on the lower empty box on the right, which counts the ways to distribute n distinct balls among k identical boxes such that every box has something in it. Alternatively, we may think of this as the number of ways to *partition* the set $[n]$ into k nonempty subsets.

A **partition of a set** A into k blocks is a set of disjoint subsets $\{A_1, A_2, \ldots, A_k\}$ such that each $a \in A$ is in exactly one A_i. For example,

$$\{\{a, d, e\}, \{c\}, \{b, f, g, h\}\}$$

is a partition of the set $\{a, b, c, d, e, f, g, h\}$ into 3 blocks.

Since the names of elements in A do not matter, we may as well let $A = [n]$ for $n \in \mathbb{N}$ and talk about "partitions of $[n]$". Our example above might then be written as

$$\{\{1, 4, 5\}, \{3\}, \{2, 6, 7, 8\}\}.$$

Note that these "partitions of $[n]$" are different from "partitions of n", since the former implies that we care about distinguishing elements of $[n]$, and the latter does not.

On the other hand, similar to partitions of n into k parts, the number of partitions of $[n]$ into k blocks does not have a nice formula. Instead, we will create a new function to count them: the **Stirling numbers of the second kind**[1] $S(n, k)$. For example, $S(4, 2) = 7$.

$$\begin{array}{cccc} \{\{1\}, \{2, 3, 4\}\} & \{\{2\}, \{1, 3, 4\}\} & \{\{3\}, \{1, 2, 4\}\} & \{\{4\}, \{1, 2, 3\}\} \\ \{\{1, 2\}, \{3, 4\}\} & \{\{1, 3\}, \{2, 4\}\} & \{\{1, 4\}, \{2, 3\}\} & \end{array}$$

[1] We will answer the obvious question later, but not now.

By analogy with binomial coefficients, another notation for $S(n,k)$ is $\left\{ n \atop k \right\}$. Some special cases of $S(n,k)$ are listed below.

$$S(n,1) = 1 \qquad\qquad S(n,n) = 1$$

$$S(n,2) = 2^{n-1} - 1 \qquad\qquad S(n,n-1) = \binom{n}{2}$$

While $S(n,k)$ has no closed-form expression, it does satisfy a recurrence relation.

Lemma 3.4.1. $S(n,k) = S(n-1,k-1) + k \cdot S(n-1,k)$

Proof. Recall that $S(n,k)$ counts partitions of the set $[n]$ into k blocks. We will break these set partitions up into cases based on what happens to the last element n. If n is in a block of its own, then we have $S(n-1,k-1)$ ways to partition $[n-1]$ into $k-1$ blocks. Otherwise, n is in a block with at least one other element; there are $S(n-1,k)$ ways to partition $[n-1]$ into k blocks, and for each partition there are k ways to add n to one of those blocks. Putting this all together, we get the recurrence relation in the lemma. $\square$

We can use this new function $S(n,k)$ to finally fill in the rest of the cells in the twelvefold way.

- *distinguishable balls, at least one per box*: We just defined $S(n,k)$ to be the formula for this problem.
- *distinguishable balls, no restriction*: We can add up $S(n,i)$ over all values $i \leq k$ to get this number.
- *distinguishable balls and boxes, at least one per box*: Since $S(n,k)$ counts distributions into identical boxes, and our number of boxes is a constant k, we can simply multiply by $k!$ to account for different permutations of boxes.

Instead of a recurrence relation, we could instead attempt to count $S(n,k)$ by counting surjective functions $f:[n]\to[k]$, then dividing by $k!$. This gives us the formula

$$S(n,k)=\sum_{i=0}^{k}(-1)^i\frac{(k-i)^n}{i!(k-i)!}\,.$$

Proof. We will use the inclusion-exclusion principle to count functions $f:[n]\to[k]$ that are *not* surjective. Let A_i denote the set of functions $f:[n]\to[k]$ such that $i\notin \operatorname{im}(f)$. The cardinality $|A_i|=(k-1)^n$, since a function that misses a particular element of the codomain is like a function $[n]\to[k-1]$. The inclusion-exclusion principle says that

$$|A_1\cup A_2\cup\cdots\cup A_k|=\sum_{i=1}^{n}(-1)^{i+1}\sum_{j_1<j_2<\cdots<j_i}|A_{j_1}\cap A_{j_2}\cap\cdots\cap A_{j_k}|$$

where the rightmost expression is the size of the intersection of i of these sets. The number of functions that miss i specified elements of the codomain is $(k-i)^n$, so we can simplify the above as

$$\sum_{i=1}^{k}(-1)^{i+1}\sum_{j_1<j_2<\cdots<j_i}(k-i)^n\,.$$

Since the inner sum is being taken over all subsets of k of size i, we can also replace it with multiplication by a binomial coefficient:

$$\sum_{i=1}^{k}(-1)^{i+1}\binom{k}{i}(k-i)^n\,.$$

This gives us a formula for functions $f:[n]\to[k]$ that are not surjective, so subtracting from the total number gives us the count of surjective ones:

$$\begin{aligned}
k^n-\sum_{i=1}^{k}(-1)^{i+1}\binom{k}{i}(k-i)^n &= k^n+\sum_{i=1}^{n}(-1)^i\binom{k}{i}(k-i)^n\\
&=(-1)^0\binom{k}{0}(k-0)^n+\sum_{i=1}^{k}(-1)^i\binom{k}{i}(k-i)^n\\
&=\sum_{i=0}^{k}(-1)^i\binom{k}{i}(k-i)^n
\end{aligned}$$

Dividing this formula by $k!$ and simplifying gives the desired result.

□

Similar to the relationship between $p_k(n)$ and $p(n)$, we may also want to count the set partitions of $[n]$ into *any* number of nonempty parts. This leads to the definition of the **Bell numbers**:

$$B(n) = \sum_{k=0}^{n} S(n,k)$$

Lemma 3.4.2. $B(n) = \displaystyle\sum_{k=0}^{n-1} \binom{n-1}{k} B(k)$

Question 3.4.3. Prove that the Bell numbers satisfy the recurrence above.

3.5 Stirling numbers of the first kind

Last time we talked about partitions of $[n]$ into k blocks, and defined the **Stirling number of the second kind** $S(n, k)$ to count these set partitions.

Today we will talk about the **(unsigned) Stirling numbers of the first kind** $c(n, k)$; we saved them for last because they don't fit into our table of the twelvefold way, but they are nevertheless interesting and perhaps surprisingly related to the other numbers we've talked about.

Here is a problem where the answer is $c(n, k)$.

Problem. You are in charge of seating n people for a fancy dinner, using k large circular tables. (The people are distinguishable, the tables are not.) Each table should seat at least one person, and theoretically has no maximum capacity. Also, we only want to consider two seating arrangements to be different if some person has a different person on their left side in each arrangement. (You could imagine the tables on rotating platforms to account for this symmetry.) How many different seating arrangements could you make?

Question 3.5.1. (a) Calculate the number of ways to seat 4 people at 2 tables as above, e.g. $c(4, 2)$.

(b) Find a recurrence relation satisfied by $c(n, k)$.

(Hint: Consider the different ways you can obtain a seating arrangement with n people and k tables by adding one person to a seating arrangement with potentially fewer people/tables.)

While the previous page gives us a nice combinatorial interpretation of $c(n,k)$, Stirling's original formulation was completely algebraic! He was interested in the numbers that you get when multiplying out rising/falling factorials, e.g.

$$\begin{aligned}
x^{\overline{0}} &= 1 && = 1 \\
x^{\overline{1}} &= x && = x \\
x^{\overline{2}} &= x(x+1) && = x^2 + x \\
x^{\overline{3}} &= x(x+1)(x+2) && = x^3 + 3x^2 + 2x \\
x^{\overline{4}} &= x(x+1)(x+2)(x+3) && = x^4 + 6x^3 + 11x^2 + 6x
\end{aligned}$$

This gives us the *algebraic definition of* $c(n,k)$, which is the coefficient of x^k in $x^{\overline{n}}$, thought of as a polynomial in x. Some people also prefer to define the **(signed) Stirling numbers of the first kind** $s(n,k)$ to be the coefficients of x^k in $x^{\underline{n}}$.

Question 3.5.2. Find a formula for the sign of $s(n,k)$.

Another interpretation of $c(n,k)$ is as the number of *permutations of* $[n]$ *with* k *cycles*. Recall that a permutation of $[n]$ is simply a bijection $\sigma : [n] \to [n]$. We can write permutations in **one-line notation** by specifying the sequence $(\sigma(1), \ldots, \sigma(n))$. A **cycle** of σ is a subset $C \subseteq [n]$ such that $\sigma\big|_C$ is a bijection $C \to C$, and C is minimal with respect to this condition. For example, the permutation $\sigma : [7] \to [7]$ defined by $(2,4,5,1,7,6,3)$ has 3 cycles: $(1,2,4)$, $(3,5,7)$, and (6).

Question 3.5.3. Find a closed-form expression for the sum

$$\sum_k c(n,k)\,.$$

3.6 Polynomial bases

Let's recall some linear algebra. A **vector space** is a set of elements called **vectors** that we can add together and scale by elements of some field; we'll assume these are real numbers, but we've seen in other examples that complex numbers would probably work too without changing anything. The usual examples of real vector spaces are $\mathbb{R}^n$ for different values of n, but another example is the set $\mathbb{P}_n$ of polynomials of degree $\leq n$.

A **basis** for a vector space V is a set of vectors that is linearly independent and spans the whole space. Every basis for a fixed V has the same cardinality; this number is called the **dimension** of V. For example, the standard basis for $\mathbb{R}^n$ is the set of standard unit vectors $\mathcal{E} = \{\vec{e_1}, \vec{e_2}, \ldots, \vec{e_n}\}$. A similar basis for $\mathbb{P}_n$ is the set $\{1, x, x^2, \ldots, x^n\}$. Note that $\mathbb{P}_n$ has one more basis vector than $\mathbb{R}^n$. Therefore, $\mathbb{P}_n$ has the same dimension as $\mathbb{R}^{n+1}$, which means these two spaces are isomorphic.

Here are some examples of bases for $\mathbb{P}_n$.

$$\begin{aligned}
&\{1, x, x^2, \ldots, x^n\} && \text{powers of } x \\
&\{1, x+1, (x+1)^2, \ldots, (x+1)^n\} && \text{powers of } x+1 \\
&\{1, x, x(x-1), \ldots, x^{\underline{n}}\} && \text{falling factorials} \\
&\{1, x, x(x+1), \ldots, x^{\overline{n}}\} && \text{rising factorials}
\end{aligned}$$

Recall that the binomial theorem tells us that

$$(x+1)^n = \sum_k \binom{n}{k} x^k \,.$$

We can interpret this equation as telling us how to convert from the "powers of $x+1$" basis to the standard "powers of x" basis. For example, here is how we might translate $(x+1)^4 - (x+1)^3 + (x+1) \in \mathbb{P}_4$ into the standard basis:

$$\begin{aligned}
&(x+1)^4 - 5(x+1)^3 + 2(x+1) \\
&= (x^4 + 4x^3 + 6x^2 + 4x + 1) - 5(x^3 + 3x^2 + 3x + 1) + 2(x+1) \\
&= x^4 - x^3 - 9x^2 - 9x - 2\,.
\end{aligned}$$

One of the major tools of linear algebra is getting matrices to do things for us. Given two bases $\mathcal{A}$ and $\mathcal{B}$ of a vector space V, we can form the **change of basis matrix** $\underset{\mathcal{A}\leftarrow\mathcal{B}}{P}$. This matrix, when applied to a vector written in the basis $\mathcal{B}$, gives us that same vector written in the basis $\mathcal{A}$. If $\mathcal{B} = \{\vec{b_1}, \ldots, \vec{b_n}\}$, then the columns of $\underset{\mathcal{A}\leftarrow\mathcal{B}}{P}$ are $[\vec{b_i}]_{\mathcal{A}}$, the elements of $\mathcal{B}$ written in the basis $\mathcal{A}$. The above example, with $\mathcal{A}$ as the powers of x and $\mathcal{B}$ as the powers of $x+1$, would correspond to the matrix multiplication:

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 3 & 6 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 2 \\ 0 \\ -5 \\ 1 \end{pmatrix} = \begin{pmatrix} -2 \\ -9 \\ -9 \\ -1 \\ 1 \end{pmatrix}.$$

Question 3.6.1. Let $\mathcal{A} = \{1, x, x^2, x^3\}$ and let $\mathcal{B} = \{1, 1+x, (1+x)^2, (1+x)^3\}$ be bases of $\mathbb{P}_3$.

(a) Find the change of basis matrix $\underset{\mathcal{A}\leftarrow\mathcal{B}}{P}$.

(b) Find the change of basis matrix $\underset{\mathcal{B}\leftarrow\mathcal{A}}{P} = \left(\underset{\mathcal{A}\leftarrow\mathcal{B}}{P}\right)^{-1}$.

Question 3.6.2. Let $\mathcal{A}$ be the basis of $\mathbb{P}_n$ consisting of powers of x, and let $\mathcal{B}$ be the basis consisting of powers of $x+1$. We know that the entry in the ith row and jth column of $\underset{\mathcal{A}\leftarrow\mathcal{B}}{P}$ is the binomial coefficient $\binom{j}{i}$. Find a formula for the entry in the ith row and jth column of $\underset{\mathcal{B}\leftarrow\mathcal{A}}{P}$, and prove it is correct.

(Hint: $x = (x+1) - 1$.)

Question 3.6.3. Let $\mathcal{A} = \{1, x, x^2, x^3\}$ as before, and let $\mathcal{C} = \{1, x, x^{\overline{2}}, x^{\overline{3}}\}$ be bases of $\mathbb{P}_3$.

(a) Find the change of basis matrix $\underset{\mathcal{A}\leftarrow\mathcal{C}}{P}$.

(b) Find the change of basis matrix $\underset{\mathcal{C}\leftarrow\mathcal{A}}{P} = \left(\underset{\mathcal{A}\leftarrow\mathcal{C}}{P}\right)^{-1}$.

Question 3.6.4. Let $\mathcal{A}$ be the basis of $\mathbb{P}_n$ consisting of powers of x, and let $\mathcal{C}$ be the basis consisting of rising factorials of x.

(a) What are the entries of $\underset{\mathcal{A}\leftarrow\mathcal{C}}{P}$?

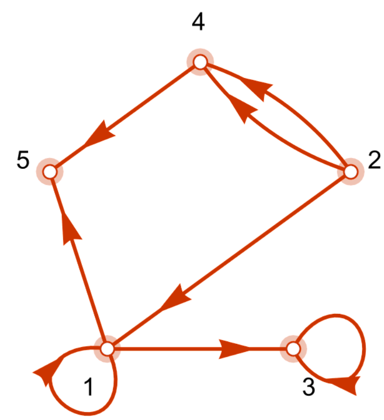
(b) What are the entries of $\underset{\mathcal{C}\leftarrow\mathcal{A}}{P}$?

Chapter 4

Graphs

4.1 Walks in graphs

For our purposes, a **(directed) graph** $G = (V, E)$ is a set V whose elements are called **vertices** and a multiset E with elements taken from $V \times V$ whose elements are called **edges**. Here's an example, with $V = [5]$ and $E = \{(1,1), (1,3), (1,5), (2,1), (2,4), (2,4), (3,3), (4,5)\}$.



A graph is called **simple** if it has no *self-edges* (v, v) like the ones at 1 and 3 above or *parallel edges* (v, w) and (v, w) like the ones from 2 to 4. If we instead define edges to be unordered sets of vertices instead of tuples, we get the definition of an **undirected graph**.

There are many different ways of representing graphs, each of which is suited for a different application. One such representation is called the **adjacency matrix**. Let $G = (V, E)$ be a graph, with $n = |V|$. The adjacency matrix $A = \mathrm{Adj}(G)$ is an $n \times n$ matrix, and is defined such that A_{ij} is the number of edges from j to i (note the reverse order). For example, the adjacency matrix of the above graph is

$$\mathrm{Adj}(G) = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

A **walk** in a graph G is a sequence of consecutive edges, i.e. a tuple $(e_1, \ldots, e_\ell)$ with the property that e_{i+1} starts where e_i ends for all i. A walk of length 1 is just an edge, so we can interpret the entries of the adjacency matrix as telling us how many walks of length 1 there are between vertices. While this rephrasing seems silly at first, it leads to a much more remarkable fact.

Theorem 4.1.1. *Let G be a graph with vertex set $[n]$, and let $A = \mathrm{Adj}(G)$. Then $(A^\ell)_{ij}$ counts the number of walks of length ℓ starting at vertex j and ending at vertex i.*

Proof. We will prove this by induction on the length of the walk ℓ.

Our base case is that $\ell = 0$. In this case, every walk of length zero is the empty set of edges, so we should expect one walk of length zero starting and ending at every vertex, and no walks of length zero connecting distinct vertices. These are exactly the entries of the $n \times n$ identity matrix I_n.

For our inductive step, we assume that $(A^{\ell-1})_{ij}$ counts walks of length $\ell - 1$ from j to i for all $i, j \in [n]$. We can use this fact to expand $(A^\ell)_{ij}$:

$$(A^\ell)_{ij} = (AA^{\ell-1})_{ij} = \sum_k A_{ik}(A^{\ell-1})_{kj}\,.$$

Note that the above expression counts pairs of walks of length $\ell - 1$ from j to any vertex k and walks of length 1 from that same vertex k to i. We can see that these pairs of walks are in bijection with walks of length ℓ from j to i, simply by concatenating the sequences of edges. Therefore, $(A^\ell)_{ij}$ counts walks of length ℓ from j to i. □

The key step in the above proof is the formula for matrix multiplication; if A is a $m \times p$ matrix and B is an $p \times n$ matrix, then the $m \times n$ matrix AB has entries given by

$$(AB)_{ij} = \sum_k A_{ik}B_{kj}\,.$$

We can use this to come up with combinatorial interpretations of matrix multiplication: $(AB)_{ij}$ counts the number of ways to combine an object counted by A_{ik} with a compatible object counted by B_{kj} over all intermediate values of k. This sum also parallels our formula for sequence convolutions; if you encounter a similar-looking sum in a different counting problem, there may be matrix multiplication involved!

An undirected graph is **connected** if every two vertices are connected by a walk. We can also use the adjacency matrix to recognize this situation.

Theorem 4.1.2. *Let G be an undirected graph with vertex set $[n]$, and let $A = \mathrm{Adj}(G)$. Then G is connected if and only if all entries of $(I_n + A)^{n-1}$ are positive.*

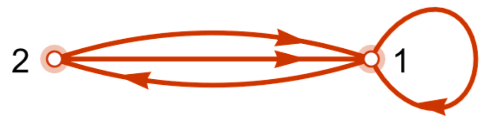
Proof. First, note that none of the entries of A are negative, so all entries are either zero or positive. We can expand the expression $(I_n + A)^{n-1}$ using the binomial theorem:

$$(I_n + A)^{n-1} = \sum_k \binom{n-1}{k} A^k\,.$$

Recall that $(A^k)_{ij}$ counts walks from j to i of length k. Therefore, if G is connected, then for any two $i, j \in [n]$, $(A^k)_{ij}$ will be positive for some k. Conversely, if $(A^k)_{ij}$ is positive for some k for any two vertices $i, j \in [n]$, then there is a walk between any two vertices, and thus G is connected. □

Question 4.1.3. In the above theorem, we restricted our scope to undirected graphs. If G is a directed graph instead, what does it mean when $(I_n + A)^{n-1}$ is positive?

Question 4.1.4. How many walks of length ℓ in the graph below start and end at vertex 1?

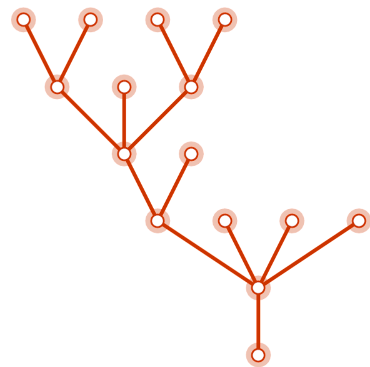


Question 4.1.5. Find a directed graph G with a vertex v that has the following property: the number of walks of length ℓ starting and ending at v is equal to the number of compositions of ℓ into any number of parts, where each part is ≤ 5.

4.2 Trees

Last time, we defined a *walk* in a graph G to be a sequence of consecutive edges. If a walk additionally does not repeat any edges, then it is called a **trail**. As an even further classification, if a trail does not repeat any vertices, it is called a **path**. If a trail starts and ends at the same vertex, but otherwise does not repeat any vertices, then it is called a **cycle**.

Graphs without cycles are a very common type of graph to study. An undirected graph without cycles is called a **tree**.



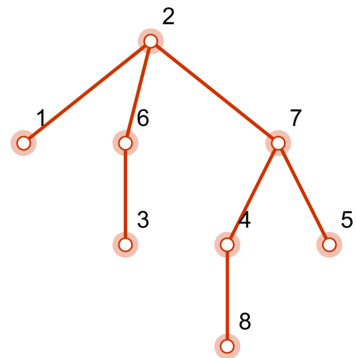
Here are some more facts about trees:

- Trees are also characterized as being *minimally connected*, i.e. they are connected, but become disconnected if any edge is removed.
- A disjoint union of trees is called a **forest**.
- A vertex in a tree that is part of only one edge is called a **leaf**. Every tree with multiple vertices has at least two leaves.
- Every tree with n vertices has $n-1$ edges.
- Trails and paths are equivalent concepts for trees.
- There is a unique trail connecting any two vertices of a given tree.

How many trees are there? If we are specifically interested in the number of distinct trees with n vertices, then it turns out that there is no nice formula or way to count these![1] However, we can make the problem much more tractable by keeping track of which vertices are which; in other words, if we assume the vertices are *distinguishable*, the problem gets a lot easier.

[1] You can find a list of the first few values at https://oeis.org/A000055.

A **labeled tree** is a tree in which each vertex is labeled with a distinct element of $[n]$, where n is the number of vertices.



Theorem 4.2.1 (Cayley's formula). *The number of distinct labeled trees on n vertices is n^{n-2}.*

There are many different proofs of this fact; we will demonstrate two of them. For our first proof, we will use a technique called **double counting**. Let T_n be the number of distinct labeled trees on n vertices. In order to find a formula for T_n using double counting, we will count some other set of objects in two different ways; one of these ways should involve T_n, and the other should not. The set of objects that we will count is the set of "directed labeled trees on n vertices with edges labeled by distinct elements of $[n-1]$".

The first way we will count these objects is by choosing some labeled tree T, then counting the number of ways to choose directions for the edges of T, and finally counting the number of ways to label the edges.

- We have T_n choices of our undirected labeled tree T.
- We have n choices for how to orient the edges of T.
- We have $(n-1)!$ choices for how to label the edges.

Therefore, the total number of directed labeled trees on n vertices with labeled edges is $T_n(n)(n-1)! = n!T_n$.

The second way that we will count these objects is to start with n isolated vertices, and count the number of ways to build up a tree by adding edges. We will label the first edge that we add by 1, the second edge by 2, and so on. For the first edge, we have n options for which vertex to end our edge at, and $n-1$ options for which vertex to start our edge at, so we have $n(n-1)$ ways to add our first edge in total. In general, if we have k disjoint trees and we want to combine two of them with an edge, we have $n(k-1)$ ways to do so. Therefore, our total count is

$$\prod_{k=2}^{n} n(k-1) = n^{n-1}\prod_{k=2}^{n}(k-1) = n^{n-1}(n-1)! = n^{n-2}n!\,.$$

Since we counted the same set of objects in two different ways, we can set our formulas equal to each other to get that

$$\begin{aligned} n!T_n &= n^{n-2}n! \\ T_n &= n^{n-2}\,. \end{aligned}$$

Let's prove Cayley's formula another way; this time, we will use **Prüfer codes**. The Prüfer code of a labeled tree T with n vertices is an element of $[n]^{n-2}$, i.e. it is a string of length $n-2$ over the alphabet $[n]$, defined as follows:

> Given a labeled tree T, we start by removing the leaf with the smallest label, and record its parent as the first element of our tuple. Then we again remove the leaf with the smallest label, and record its parent as the second element of our tuple. We repeat this process of removing the smallest leaf and recording its parent until we are left with only two vertices in our tree, at which point we stop.

For example, the Prüfer code of the tree from the previous page is $(2, 6, 7, 2, 7, 4)$. Note that we stop when we are left with two vertices because there is a unique labeled tree on two vertices, so we would not be recording any new information by continuing.

The inverse to this process, turning a Prüfer code back into a labeled tree, also has a nice description:

> Let P be our Prüfer code. Start with n isolated vertices, which we will eventually connect to form our tree. Also, separately keep track of a list L of the numbers $1, \ldots, n$. Find the smallest number v that is in L but not in P, and connect vertex v to the first element of P. Remove v from L and remove the first element of P, then repeat this process until our Prüfer code is empty. Finally, connect the two remaining vertices in L to complete our tree.

Question 4.2.2. How many labeled trees on n vertices are there in which …

(a) …the vertex 1 is a leaf?

(b) …the vertex 1 is adjacent to exactly two other vertices?

(c) …exactly two vertices are leaves?

4.3 Random walks in $\mathbb{Z}^d$

We have previously discussed walks on finite graphs; today we will look at a rather famous result regarding walks on the infinite graph $\mathbb{Z}^d$. If $d \in \mathbb{N}$, then we define the undirected simple graph $\mathbb{Z}^d$ such that the vertices are d-tuples of integers and the edges connect two d-tuples that differ by 1 in exactly one coordinate. For example, the lattice paths that we have previously studied are examples of paths in $\mathbb{Z}^2$. We can think of a walk on $\mathbb{Z}^d$ as a sequence of steps of length 1 in one of the $2d$ cardinal directions. A **random walk** on $\mathbb{Z}^d$ is an infinitely long walk, where each step is chosen (uniformly) randomly from the $2d$ options. Our main question today concerns such random walks:

Problem 4.3.1. What is the probability that an infinite random walk starting at the origin in $\mathbb{Z}^d$ eventually returns to the origin?

More specifically, we will approximate the expected number of times a random walk in $\mathbb{Z}^d$ returns to the origin, given infinite time. If the expected number of returns to the origin is ∞, then one can show that on average every random walk returns to the origin eventually. On the other hand, if the expected number of returns is a finite number, then on average many random walks must never return to the origin.

One dimension

Let's start off with the easiest case of random walks on $\mathbb{Z}^1$. In order to estimate the expected number of times a walk returns to the origin, we will fix some length $2n$ and count the probability $p(2n)$ that the random walk returns to the origin on step $2n$. The expected number of returns to the origin is then the sum of $p(2n)$ over all values of $n \in \mathbb{N}$, i.e.

$$E = \sum_{n=1}^{\infty} p(2n)\,.$$

To calculate $p(2n)$, we can count all the walks of length $2n$ that return to the origin on step $2n$. We can model a walk of length $2n$ as a string of $2n$ symbols chosen from $\{L, R\}$ corresponding to steps to the left and right, respectively. A walk of length $2n$ that returns to the origin on step $2n$ is therefore a string of exactly n Ls and n Rs. There are 2^{2n} total strings of Ls and Rs, and $\binom{2n}{n}$ of them have an equal number of Ls and Rs, so we get that

$$p(2n) = \frac{\binom{2n}{n}}{2^{2n}}\,.$$

While there are methods of figuring out the sum of all values of $p(2n)$ exactly, we are only interested in a rough approximation, so we will "simplify" this expression using Stirling's formula:

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n\,.$$

Plugging this into our equation for $p(2n)$ gives us

$$p(2n) \approx \frac{1}{2^{2n}}\frac{\sqrt{2\pi(2n)}(\frac{2n}{e})^{2n}}{\left(\sqrt{2\pi(n)}(\frac{n}{e})^n\right)^2} = \frac{1}{2^{2n}}\frac{\sqrt{2}\cdot 2^{2n}}{\sqrt{2\pi n}} = \frac{1}{\sqrt{\pi n}}\,.$$

Therefore, our expected number of returns is

$$E = \sum_{n=1}^{\infty} p(2n) \approx \sum_{n=1}^{\infty} \frac{1}{\sqrt{\pi n}} = \infty\,.$$

Two dimensions

We will calculate the same expected number of returns E and probability of return $p(2n)$ as before, this time in $\mathbb{Z}^2$. The only difference is that we now model a random walk as a string of $2n$ symbols from $\{L, R, U, D\}$, since we can now step left, right, up, and down. There are 4^{2n} such strings, and we can calculate the number of strings with an equal number of Ls as Rs and Us as Ds to be

$$\sum_{k=0}^{n} \binom{2n}{k, k, n-k, n-k}$$

which means that our probability of return at step $2n$ is

$$p(2n) = \frac{1}{4^{2n}} \sum_{k=0}^{n} \binom{2n}{k, k, n-k, n-k}.$$

Again, we can use Stirling's approximation (but skip the algebra) to get that

$$p(2n) \approx \frac{1}{\pi n}$$

so our expected number of returns is

$$E = \sum_{n=1}^{\infty} p(2n) \approx \sum_{n=1}^{\infty} \frac{1}{\pi n} = \infty\,.$$

Three dimensions

Finally, we get to the interesting case! We repeat all the same steps as before, but this time our third dimension means that we are talking about walks in $\mathbb{Z}^3$, which we can model as strings on the alphabet $\{L, R, U, D, F, B\}$, where the extra letters correspond to e.g. "forward/backward" or "front/back". Our probability of return at step $2n$ is

$$p(2n) = \frac{1}{6^{2n}} \sum_{k=0}^{n} \sum_{\ell=0}^{k} \binom{2n}{k, k, \ell, \ell, n-k-\ell, n-k-\ell},$$

which we use Stirling's formula to approximate as

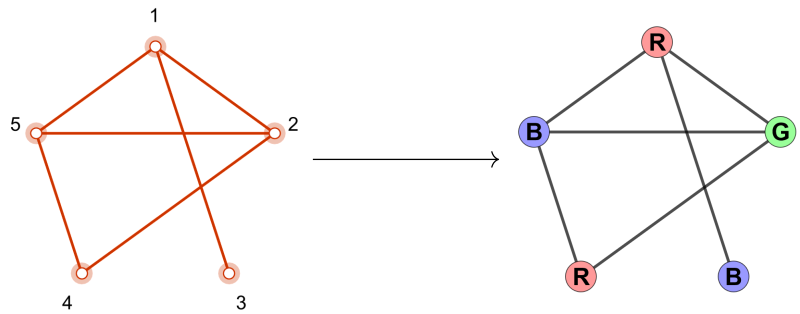
$$p(2n) \approx \frac{1}{(\pi n)^{3/2}}\,,$$

giving us an expected number of returns of

$$E \approx \sum_{n=1}^{\infty} \frac{1}{(\pi n)^{3/2}} \approx 0.468 < \infty\,.$$

4.4 Graph coloring

A k**-coloring** of an undirected graph $G = (V, E)$ is a function $f : V \to [k]$ with the property that $f(v) \neq f(w)$ for all edges $\{v, w\} \in E$. Here, we think of $[k]$ as our set of "colors"; the condition that f satisfies can then be stated as "no two adjacent vertices should be painted the same color". Here is an example graph G, along with a 3-coloring.



Given a graph G, how many different k-colorings are there? Whatever the answer is, we will denote this number by $\chi_G(k)$. Let's start small and let G be the path graph P_5 with 5 edges and 6 vertices depicted below.

Some examples of $\chi_G(k)$ for $k = 1, 2, 3$:

- We can clearly not color G with a single color, so $\chi_G(1) = 0$.
- On the other hand, if we have two colors, then we can alternate them to come up with a valid 2-coloring of G; furthermore, we have two options for our starting color, so $\chi_G(2) = 2$.
- If $k = 3$, then we can choose any of our three colors to start with, and after that we have two options for what color to choose next, giving us $3 \cdot 2^5 = 96$ total options.

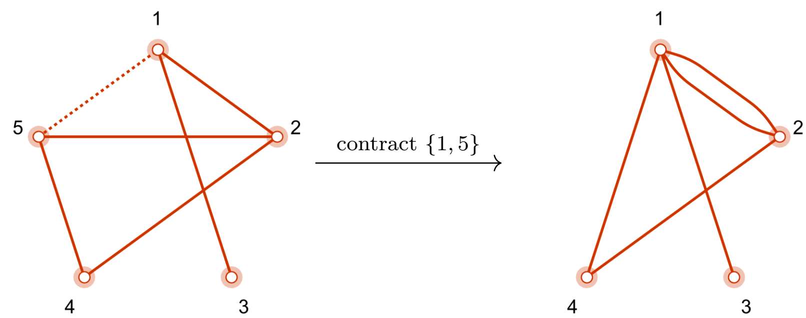
In general, if we have k colors, then our 3-coloring logic tells us that $\chi_G(k) = k(k-1)^5$.

Question 4.4.1. If $G = K_n$, the complete graph on n vertices, find a simple formula for $\chi_G(k)$.

It turns out that, for a fixed graph G, $\chi_G(k)$ will always be a polynomial in k, which is why it is known as the **chromatic polynomial**. How can we prove this? Induction seems like a likely strategy, so we will first need to find a recurrence relation that lets us break down our graphs into smaller ones.

Consider a graph $G = (V, E)$ that we would like to count the k-colorings of, and choose some edge $e = \{v, w\}$. We can consider the graph $G \setminus e$ where we remove the edge e but keep all the vertices. Almost all k-colorings of $G \setminus e$ are also valid k-colorings of G; the only colorings that are not valid are those that assign v and w the same color. Therefore, we can calculate $\chi_G(k)$ from $\chi_{G\setminus e}(k)$ by subtracting out these invalid colorings. But how do we count such colorings?

Given an edge e in a graph $G = (V, E)$, we can get a new graph G/e by **contracting** the edge. This operation produces a graph with vertex set $V \setminus w$, and edge set $E \setminus e$ with two changes; any edge previously containing w now contains v instead. Visually, G/e is obtained by shrinking e so that the vertices v and w "merge" into a single vertex, which we just call v for convenience.



Note that, even when G is simple, $G \setminus e$ might end up with self-edges and/or parallel edges.

This operation is helpful to us because a k-coloring of G/e is exactly a k-coloring of $G \setminus e$ where v and w are painted the same color. Therefore, if we want to count k-colorings of G, we can first over-count by considering k-colorings of $G \setminus e$, then subtract k-colorings of G/e. This leads us to the **deletion-contraction recurrence** for counting k-colorings of G.

Theorem 4.4.2 (The deletion-contraction recurrence)**.** $\chi_G(k) = \chi_{G\setminus e}(k) - \chi_{G/e}(k)$.

Question 4.4.3. Use the deletion-contraction recurrence to compute $\chi_G(k)$ for the graph G on the first page.

Now that we have the deletion-contraction recurrence, we can prove that $\chi_G(k)$ is always a polynomial.

Proof that $\chi_G(k)$ is a polynomial in k for every graph G. We will proceed by induction on the number of edges m in G. For our base case, assume that $m = 0$, i.e. G has no edges. Let n be the number of vertices in G. We can pick any of the k colors for each vertex, so $\chi_G(k) = k^n$, which is a polynomial in k.

For the inductive step, assume that G is a graph with $m \geq 1$ edges. Assume that $\chi_G(k)$ is a polynomial for all graphs G with $< m$ edges. The deletion-contraction recurrence tells us that

$$\chi_G(k) = \chi_{G \setminus e}(k) - \chi_{G/e}(k)\,.$$

Since $G \setminus e$ and G/e are graphs with $m - 1$ edges, $\chi_{G\setminus e}(k)$ and $\chi_{G/e}(k)$ are polynomials. Therefore, $\chi_G(k)$ is a difference of polynomials, and thus also a polynomial. $\square$

Question 4.4.4. Let $e \in K_5$ be any edge; find $\chi_{K_5 \setminus e}(k)$.

Question 4.4.5. It turns out that all trees with n vertices has the same chromatic polynomial. Prove this fact by finding a formula for $\chi_T(k)$, where T is any tree with n vertices.

(Hint: Trees have leaves.)

Question 4.4.6. Let C_n be the cycle graph with n edges and n vertices arranged in a circle. Find a formula for $\chi_{C_n}(k)$.

Question 4.4.7. Recall that a **spanning tree** of a graph G is a subgraph $T \subseteq G$ that contains all the vertices of G and is also a tree. Let $t(G)$ denote the number of distinct spanning trees in the graph G. Find a recurrence relation satisfied by $t(G)$ and use it to compute the number of spanning trees of the graph on the first page.

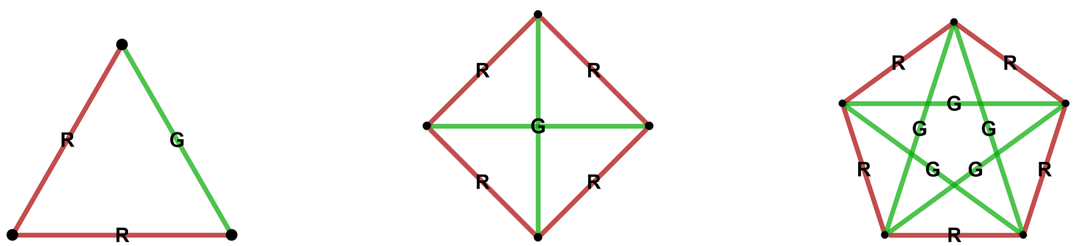
4.5 Ramsey theory

Suppose there are n people at a party, and you want to find either three people who all know each other, or three people who all do not know each other. For what values of n is this always possible?

This is a classic problem in Ramsey theory, and one solution involves turning it into a graph coloring problem. Unlike last time, today we are coloring the *edges* of the graph, and also unlike vertex colorings, we will not require any restrictions on our edge colorings. A k-coloring of the edges of a graph $G = (V, E)$ will simply be a function $f : E \to [k]$.

How does this apply to our original problem? We can represent the relationships between the n people at the party as a coloring of the complete graph K_n. Each vertex represents a person. If two people know each other, we color the edge between them green; otherwise, we color the edge red. Our problem is therefore asking for what values of n we can always find a monochromatic clique of size 3, i.e. a subgraph isomorphic to K_3 with all edges colored the same.

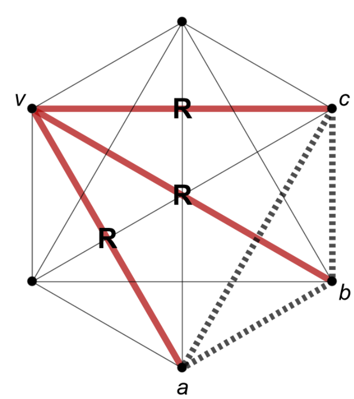
This is clearly not possible for $n \leq 2$, and we can find counter-examples for $n \in \{3, 4, 5\}$, shown below.



It turns out that, when $n = 6$, any edge 2-coloring of K_6 will contain a monochromatic K_3. We could prove this by simply drawing every edge 2-coloring of K_6, however, there are 32 768 of these, so I would rather not. Here's a nicer proof.

Proof that every edge 2-coloring of K_6 contains a monochromatic K_3. Fix a vertex $v \in K_6$, and consider the five edges that contain v. Since edges can be one of two colors, by the generalized pigeonhole principle, three of these edges must be the same color; without loss of generality, assume that at least three edges containing v are red.

Let $a, b, c \in K_6$ be the vertices at the other endpoints of these three edges, and consider the colors of the edges $\{a, b\}$, $\{b, c\}$, and $\{a, c\}$. If any of these edges are red, then it completes a red triangle with v; otherwise, these three edges themselves make up a green triangle. Therefore, any edge 2-coloring of K_6 contains a monochromatic triangle.



□

Our proof on the previous page establishes that the *Ramsey number* $R(3,3) = 6$. The **Ramsey number** $R(p,q)$ is defined to be the smallest n such that any edge 2-coloring of the edges of K_n contains either a red K_p or a green K_q. **Ramsey theory** is the subfield of combinatorics that concerns these numbers, as well as related ideas; the general theme is that of determining how "big" a structure needs to be in order to satisfy a property.

Note. Our definition of Ramsey numbers is valid, but it is not immediately clear that $R(p,q)$ should always be well-defined; what if it were possible to color the edges of K_n for all n without monochromatic cliques of some size?

Let's figure out some values of $R(p,q)$ for small $p,q \in \mathbb{N}$. If $q = 1$, for example, $R(p,1)$ asks us to find the smallest n for which every edge 2-coloring of K_n has either a red K_p or a green K_1, which is... a single vertex. Therefore, $R(p,1) = 1$ for all $p \in \mathbb{N}$.

If $q = 2$, then to find $R(p,2)$, we need to find the smallest n for which every edge 2-coloring of K_n has either a red K_p or a green K_2. Since a green K_2 is simply a green edge, this condition is satisfied unless we only color the edges red. Therefore, if we color all edges red, then we can avoid creating a red K_p for all $n < p$, but are forced to color K_p red when $p = n$. This means that $R(p,2) = p$ for all $p \in \mathbb{N}$.

One tool to help us compute $R(p,q)$ for larger values of $p,q \in \mathbb{N}$ is the recursive inequality

$$R(p,q) \leq R(p-1,q) + R(p,q-1)\,,$$

the proof of which is on the next page. Since this inequality relates $R(p,q)$ to previous values in the sequence, we can also use it to prove that $R(p,q)$ exists for any $p,q \in \mathbb{N}$. The inequality above is our inductive step, and our base case can consist of the values we found above: $R(p,1) = R(1,q) = 1$ for all $p,q \in \mathbb{N}$.

Question 4.5.1. Our inequality above tells us that $R(4,3) \leq R(3,3) + R(4,2) = 6 + 4 = 10$.

(a) Find examples of edge 2-colorings of K_n with no red K_4 or green K_3 for small n $(6 \leq n \leq 10)$ to establish a lower bound for $R(4,3)$.

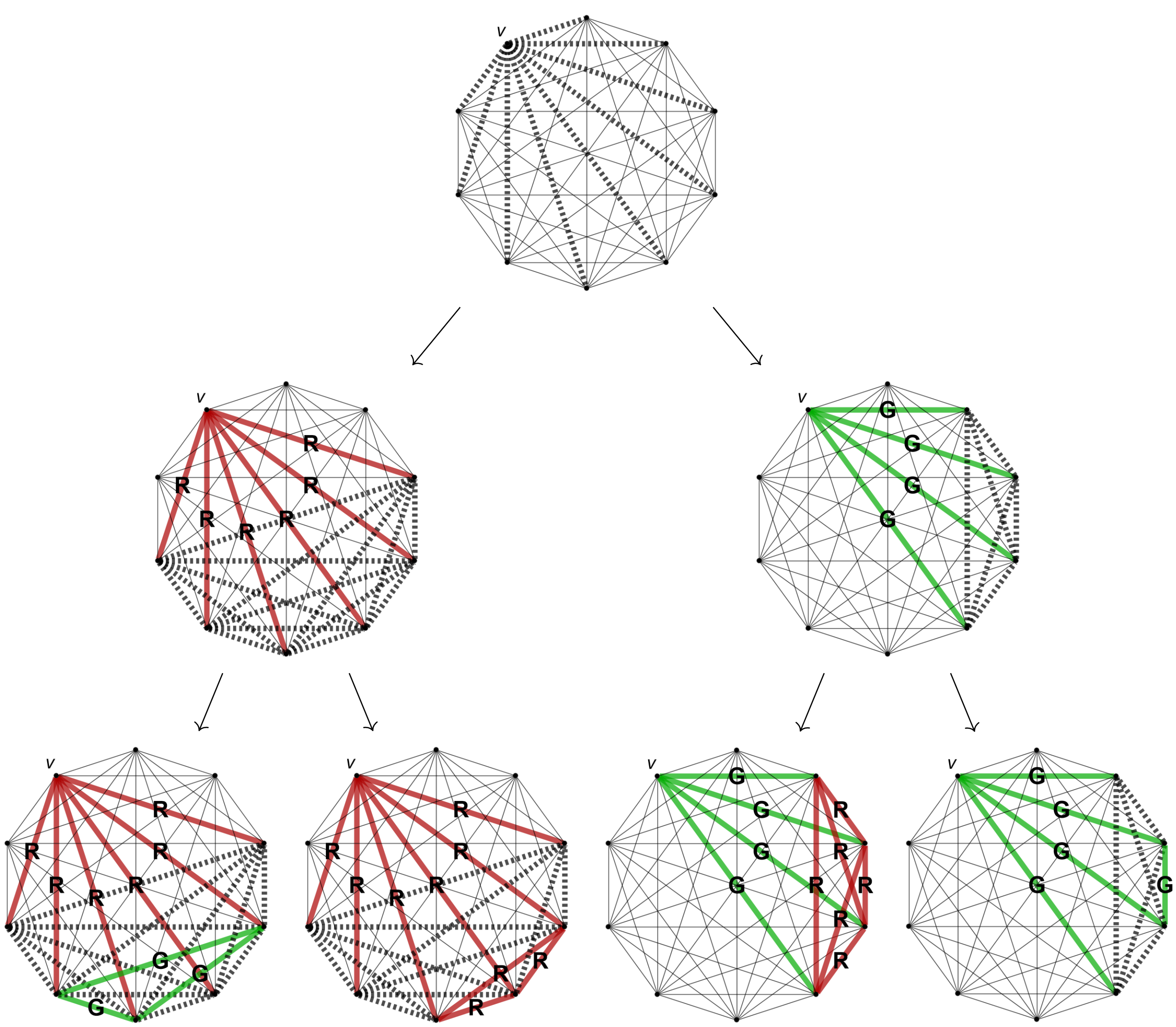
(b) Make a conjecture about the value of $R(4,3)$.

Proof. Consider the complete graph on $n = R(p-1,q) + R(p,q-1)$ vertices, and choose any edge 2-coloring. We would like to show that K_n has either a red K_p or a green K_q subgraph. Fix any vertex $v \in K_n$; by similar logic as our $K(3,3)$ proof, v is part of at least $R(p-1,q)$ red edges, or $R(p,q-1)$ green edges.

Assume the first case is true, that v is adjacent to at least $R(p-1,q)$ red edges. Consider the complete subgraph induced by the other endpoints of those red edges. By the definition of $R(p-1,q)$, this subgraph contains either a red K_{p-1} or a green K_q. In the latter case, we have found a green K_q in our graph and are done. Otherwise, we have found that the vertex v is connected to a red K_{p-1} by entirely red edges, and therefore have found a red K_p in our graph, which also means we are done.

The second case works out almost identically. If v is adjacent to $R(p,q-1)$ green edges, then we consider the subgraph induced by the other endpoints of these edges and either find a red K_p or a green K_{q-1} that we can complete to a green K_q by adding v. □

Here's an illustration of our proof that $R(4,3) \leq 10$.



Proof that $R(4,3) = 9$. We have previously found examples to show that $9 \leq R(4,3) \leq 10$, so it suffices to prove that any edge 2-coloring of K_9 must contain a red K_4 or a green K_3.

Let $v \in K_9$ be any vertex. By our logic from before, if v is adjacent to at least 6 red edges or 4 green edges, then we can find a red K_4 or a green K_3 by considering the endpoints of those edges. However, unlike before, the pigeonhole principle does not guarantee that this is always the case. We now have a third case to consider: that v is adjacent to exactly 5 red edges and 3 green edges.

Assume we are in this third case. Since v was an arbitrary choice, we can actually assume that every vertex is adjacent to exactly 5 red edges and 3 green edges. If we consider now the subgraph $G \subset K_4$ consisting of only green edges, we see that G must have 9 vertices, each of which has degree 3. However, this is impossible; therefore, every edge 2-coloring of K_9 must have a red K_4 or a green K_3. □

While we have been able to calculate $R(p,q)$ for some small values of $p, q \in \mathbb{N}$, it is in general a *very* difficult problem to find $R(p,q)$. Here is a short list of every known Ramsey number.

$R(p,1) = 1$	$R(p,2) = p$	$R(3,3) = 6$	$R(4,4) = 18$
		$R(4,3) = 9$	$R(5,4) = 25$
		$R(5,3) = 14$	
		$R(6,3) = 18$	
		$R(7,3) = 23$	
		$R(8,3) = 28$	
		$R(9,3) = 36$	

We are able to establish decent bounds for many $R(p,q)$ not on this list; for example, it is known that $43 \leq R(5,5) \leq 48$. However, a brute-force approach to checking the exact value of $R(5,5)$ would take a *very* long time; there are 6762169998536515153309949246931412563441245773262355483237897075541425952726078201272540875362012005051832255913691247089694048761634374876806898924325626584427349555187265077359763426258258445478710181225103211573094762147219990257131480304218066899066093835491046378 7008 possible edge 2-colorings of K_{43} to consider!

Question 4.5.2. Prove that $R(p,q) \leq \binom{p+q-2}{p-1}$.

Chapter 5

Permutations

5.1 Pattern-avoiding permutations

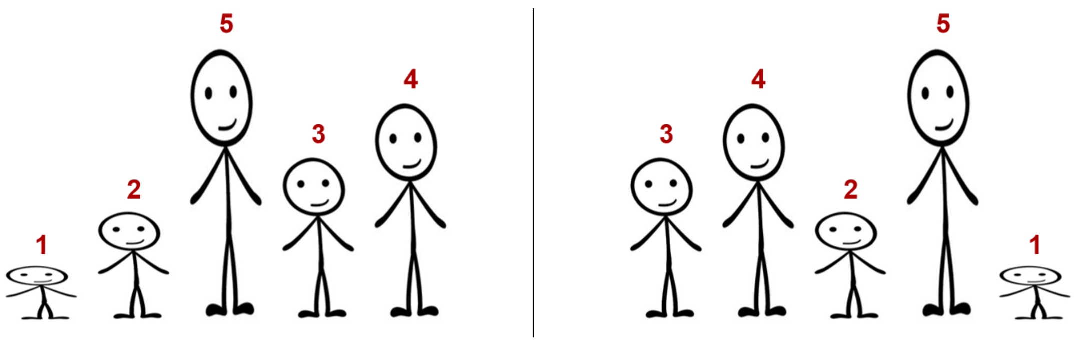
Previously, we have discussed *permutations*, i.e. bijections $\sigma : [n] \to [n]$, and counted that there were $n!$ of them. We have also counted special subsets of permutations, for example

- *derangements*, which we used the inclusion-exclusion principle to enumerate, and
- permutations with k *cycles*, which we related to the Stirling numbers of the first kind $c(n, k)$.

Today we'll be looking at *patterns* in permutations. Let's start with the following counting problem:

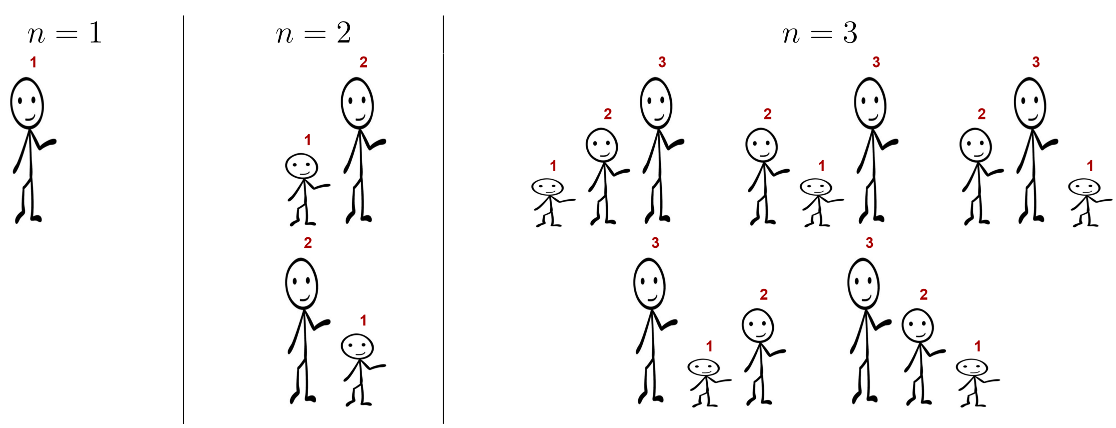
> "Let us assume that there are n children playing in our backyard, no two of whom have the same height. For the next game, they need to stand in a line so that everyone faces the back of the preceding person. Moreover, each child must be able to see all children that are shorter than them and precede them in the line. How many such lineups exist?"

For example, here are two possible lines of $n = 5$ children, named based on their relative heights (e.g. child 2 is shorter than child 3). Note that the children are looking to the left, so we want to make sure that their view of any shorter children to their left is not blocked.

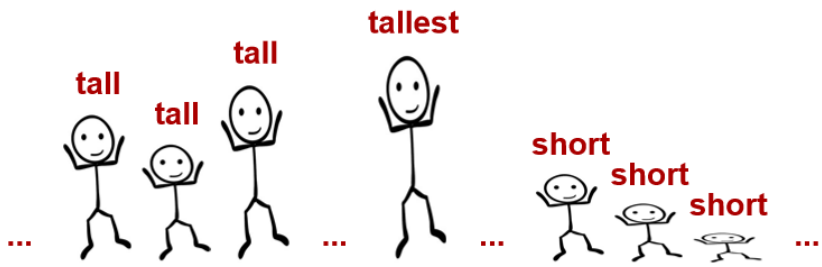


The line on the left fails this condition for multiple reasons, one of which is that the child of height 3 cannot see the child of height 1, despite being taller than them and placed after them in the line. On the other hand, the line on the left satisfies this condition; every child can see all shorter children in front of them.

How many distinct lines can we form out of n children this way? Let's try some small examples.



Let a_n denote the number of lines of n children satisfying this property. Our work on the previous page shows that $a_1 = 1$, $a_2 = 2$, and $a_3 = 5$. We can also assume $a_0 = 1$ for trivial reasons. Let's try to find a recurrence relation satisfied by a_n. Given some line of children $(c_1, \dots, c_n)$, let k be the index of the tallest child, i.e. $k \in \mathbb{N}$ is chosen such that $c_k = n$. Since the tallest child blocks the view of every child after them, it must be true that every child after them is shorter than every child before them.



Therefore, we don't need to consider any other comparisons between children on both sides of the tallest child. In other words, we can treat each side of the tallest child as its own smaller instance of the same problem. If the tallest child is at position k, then there are a_{k-1} ways to arrange the children before them, and a_{n-k} ways to arrange the children after them. Since the tallest child can be at any position, we can sum over all values of k to get that

$$a_n = \sum_{k=1}^{n} a_{k-1} a_{n-k} \, .$$

This recurrence relation, along with our initial condition $a_0 = 1$, uniquely specifies our sequence a_n. Furthermore, this is actually a sequence that we have seen before; a_n is identical to the sequence of Catalan numbers C_n, which we defined via the recurrence

$$C_n = \sum_{k=0}^{n-1} C_k C_{n-k-1} \, .$$

Since we previously obtained the closed-form expression $C_n = \frac{1}{n+1}\binom{2n}{n}$, we know that this formula also works for our identical sequence a_n.

Our problem above can be described succinctly as counting permutations of $[n]$ that avoid the pattern 132. Let $\rho = (\rho_1, \dots, \rho_k)$ be a permutation of $[k]$; a **ρ-pattern** in a permutation $\sigma = (\sigma_1, \dots, \sigma_n)$ is a subsequence $(\sigma_{i_1}, \dots, \sigma_{i_k})$ that has the same length and "shape" as ρ, i.e. $\sigma_{i_a} < \sigma_{i_b} \iff \rho_a < \rho_b$ for all indices a, b. A **ρ-avoiding permutation** σ is one that contains no ρ-patterns. We might use the notation $S_n(\rho)$ to denote the set of permutations $\sigma \in S_n$ that avoid the pattern ρ.

Therefore, our lines of children avoided the pattern 132 because they never contained a subsequence of three children with the shortest child first, the tallest child second, and the middlest child third.

Question 5.1.1. (a) How many 312-avoiding permutations of length n are there?

(b) For which other patterns ρ can you count $S_n(\rho)$?

It seems more difficult to count 123-avoiding permutations directly using a recurrence, so here's another method: a bijection $f : S_n(132) \to S_n(123)$. The bijection f takes $\sigma \in S_n(132)$ and constructs a permutation that avoids 123 as follows:

- Find all the *left-to-right minima* of σ, i.e. all the numbers that are smaller than every number before them.
- Keep the left-to-right minima in the same positions, and sort all the other numbers around them in decreasing order.

For example, given the permutation 456123, we find that the left-to-right minima are 4 and 1, so we fix these numbers in their original positions: $4 - -1 - -$. We then fill in the rest of the numbers in decreasing order to get that $f(456123) = 465132$.

Question 5.1.2. Why is the image of σ under f always a permutation in $S_n(123)$?

It turns out that the number of permutations avoiding ρ is the same for each $\rho \in S_3$.

Theorem 5.1.3. *For any $\rho \in S_3$, $|S_n(\rho)| = C_n$.*

This is not true for larger patterns; for example, here are the sizes of $S_n(\rho)$ for some particular choices of $\rho \in S_n$:

$$\begin{aligned}
|S_n(1342)| &= 1, 2, 6, 23, 103, 512, 2740, 15485, \ldots \\
|S_n(1234)| &= 1, 2, 6, 23, 103, 513, 2761, 15767, \ldots \\
|S_n(1324)| &= 1, 2, 6, 23, 103, 513, 2762, 15793, \ldots .
\end{aligned}$$

Much is still unknown about pattern avoidance in permutations, and it is an active area of research in modern combinatorics.

Let's end with an interesting application of pattern avoidance in computer science: *stack sortable permutations*. Our goal is to sort a permutation of n into the ordered list of numbers $1, \ldots, n$, but the catch is that we are only fed one number at a time. Our only available operations that we can do with our numbers are

- to add the current number to the end of the output list,
- to add the current number to top of a *stack*, i.e. a vertical array of numbers that only allows access to the topmost element, or
- to remove the top number of the stack and add it to the end of the output list.

It turns out that only some permutations of n are able to be fully sorted with these limited operations; we call these permutations **stack sortable**. These permutations also have another characterization in terms of pattern avoidance.

Theorem 5.1.4. *A permutation is stack sortable if and only if it is 231-avoiding.*

Similarly, we could replace our stack with a *queue*, which only allows us to remove numbers in the order they were put in. This would give us the notion of a *queue sortable permutation*.

Theorem 5.1.5. *A permutation is queue sortable if and only if it is 321-avoiding.*

We could also generalize the problem by allowing multiple stacks or queues. For example, a *two-stack sortable permutation* is one that is able to be sorted in two rounds through our stack-sorting process.

Theorem 5.1.6. *A permutation is two-stack sortable if and only if it does not contain a 2341-pattern, and it does not contain a 3241-pattern, except as a part of a 35241-pattern.*

While we won't prove it, here's a closed-form expression for the number of two-stack sortable permutations of $[n]$:

$$\frac{2}{(n+1)(2n+1)}\binom{3n}{n}.$$

5.2 Exponential generating functions

Recall that a *generating function* is a way of representing a sequence as a formal power series. We have previously studied *ordinary generating functions*, which encode a sequence (a_n) as the formal power series

$$A(x) = \sum_{n=0}^{\infty} a_n x^n \,.$$

We have already seen that this correspondence is very powerful and allows us to modify and combine sequences in interesting ways. One downside of OGFs, however, is that they are not ideal for working with fast-growing sequences. For example, if $a_n = n!$ is the sequence that counts permutations of $[n]$, then we can consider the OGF

$$A(x) = 1 + x + 2x^2 + 6x^3 + 24x^4 + \dots \,.$$

While this is a perfectly valid *formal* power series, it does not represent a real-valued function on any domain other than $\{0\}$, i.e. the sum is guaranteed to diverge for any $x \neq 0$. This limits how we can manipulate and combine such generating functions.

One solution to this issue is to deal instead with *exponential generating functions*. The **exponential generating function** $\mathcal{A}(x)$ of a sequence (a_n) is defined to be the formal power series

$$\mathcal{A}(x) = \sum_{n=0}^{\infty} a_n \frac{x^n}{n!} \,.$$

Here are the OGFs and EGFs of some common sequences.

Sequence	Terms	OGF	EGF
$a_n = 1$	$1, 1, 1, 1, \dots$	$\dfrac{1}{1-x}$	e^x
$a_n = r^n$	$1, r, r^2, r^3, \dots$	$\dfrac{1}{1-rx}$	e^{rx}
$a_n = n$	$0, 1, 2, 3, \dots$	$\dfrac{x}{(1-x)^2}$	xe^x
$a_n = \binom{m}{n}$	$\binom{m}{0}, \binom{m}{1}, \dots$	$(1+x)^m$	
$a_n = \binom{n}{k}$	$\binom{0}{k}, \binom{1}{k}, \dots$	$\dfrac{x^k}{(1-x)^{k+1}}$	$\dfrac{x^k}{k!}e^x$
$a_n = \dfrac{1+(-1)^n}{2}$	$1, 0, 1, 0, \dots$	$\dfrac{1}{1-x^2}$	$\dfrac{1}{2}(e^x + e^{-x})$
$a_n = n!$	$1, 1, 2, 6, \dots$		$\dfrac{1}{1-x}$

We can also represent some sequence operations as operations on EGFs, and vice versa. However, many of the operations on OGFs and EGFs are not the same!

Sequences		OGFs	EGFs
Operation	Formula	Formula	Formula
sum	$a_n + b_n$	$A(x) + B(x)$	$\mathcal{A}(x) + \mathcal{B}(x)$
convolution	$\sum_k a_k b_{n-k}$	$A(x)B(x)$	–
binomial convolution	$\sum_k \binom{n}{k} a_k b_{n-k}$	–	$\mathcal{A}(x)\mathcal{B}(x)$
right shift	a_{n-1}	$xA(x)$	$\int_0^x \mathcal{A}(t)\,\mathrm{d}t$
left shift	a_{n+1}	$\frac{A(x) - a_0}{x}$	$\mathcal{A}'(x)$
partial sum	$\sum_{k \le n} a_k$	$\frac{A(x)}{1-x}$	–
binomial sum	$\sum_{k \le n} \binom{n}{k} a_k$	–	$e^x \mathcal{A}(x)$
index multiply	$(n+1)a_{n+1}$	$A'(x)$	–
	na_{n-1}	–	$x\mathcal{A}(x)$
index divide	$\frac{a_{n-1}}{n}$	$\int_0^x A(x)\,\mathrm{d}x$	–
	$\frac{a_{n+1}}{n+1}$	–	$\frac{\mathcal{A}(x) - a_0}{x}$
–	$c^n a_n$	$A(cx)$	$\mathcal{A}(cx)$
–	$a_n - a_{n-1}$	$(1-x)A(x)$	$\mathcal{A}(x) - \int_0^x \mathcal{A}(t)\,\mathrm{d}t$

Aside from the aforementioned convergence considerations, there is a more conceptual distinction between OGFs and EGFs.

Idea. Ordinary generating functions tend to be suitable for counting indistinguishable objects, whereas exponential generating functions are usually better for counting distinct objects.

One way to see this difference is to consider the same formal power series as both an OGF and an EGF for different sequences.

Formal Power Series	Type	Sequence	Interpretation
$\dfrac{1}{1-x}$	Ordinary	$a_n = 1$	Ways to order n identical objects
	Exponential	$b_n = n!$	Ways to order n distinct objects
$\dfrac{1}{(1-x)^2}$	Ordinary	$a_n = n+1$	Ways to split n identical objects into two piles
	Exponential	$b_n = (n+1)n!$	Ways to split n distinct objects into two ordered lines
$\dfrac{1}{1-x-x^2}$	Ordinary	$a_n = F_{n+1}$	Ways to cover a $1 \times n$ grid by single tiles and dominoes.
	Exponential	$b_n = n!F_{n+1}$	Ways to cover a $1 \times n$ grid by single tiles and dominoes, where each grid cell is labeled by a distinct number in $[n]$.

Question 5.2.1. (a) Interpret the formal power series $\frac{1}{1-2x}$ as the OGF of a sequence (a_n) and give an example of a combinatorial interpretation of (a_n).

(b) Interpret the same formal power series $\frac{1}{1-2x}$ as the EGF of a sequence (b_n) and give an example of a combinatorial interpretation of (b_n).

Exponential generating functions are particularly suitable for counting *functions*, since we usually consider the domain and codomain to be sets of distinct elements.

Type of Function $[n] \to [k]$	Exponential Generating Function
k fixed, no restriction	$(e^x)^k$
k fixed, injective	$(1+x)^k$
k fixed, surjective	$(e^x - 1)^k$
k fixed, bijective ($n = k$)	$\dfrac{1}{1-x}$

Question 5.2.2. Find the exponential generating function of each sequence.

(a) $a_n = \sum_k \binom{n}{k} 2^k$.

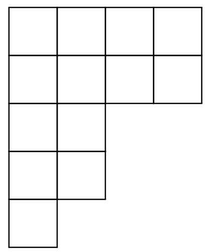
(b) $b_n = (n-1)!\,$.

(c) $c_n = n^2$.

5.3 Young tableaux

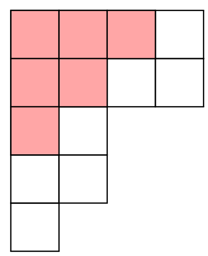
A partial order on partitions

Recall that a *partition* of n is a multiset of natural numbers that adds up to n. While partitions are unordered, we usually sort them in decreasing order to make it easy to compare them. For example, $\lambda = (4, 4, 2, 2, 1)$ is a partition of 13. We can represent partitions using *Young diagrams*, like the one below.



One feature of partitions that we have not discussed so far is that they have a natural *partial order*. Given any two partitions λ and μ, we say that $\lambda \leq \mu$ if $\lambda_i \leq \mu_i$ for all i. Visually, $\lambda \leq \mu$ if the Young diagram of λ fits inside the diagram of μ.

$(3, 2, 1) \leq (4, 4, 2, 2, 1)$



This is called a **partial order** because it is a binary relation that is reflexive, antisymmetric, and transitive, i.e. for all partitions λ, μ, ν:

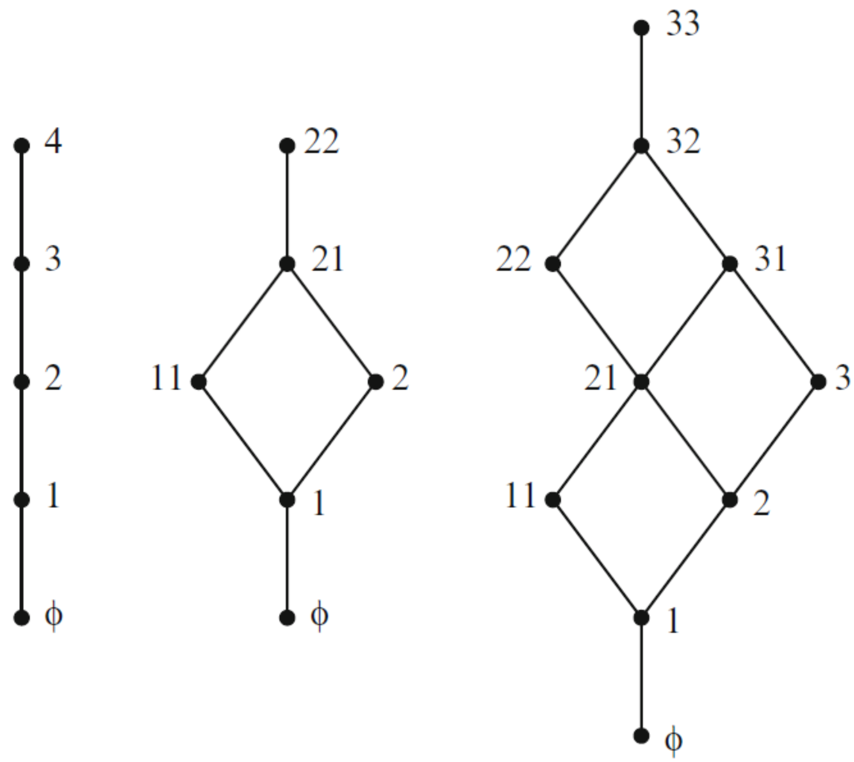
- $\lambda \leq \lambda$,
- if $\lambda \leq \mu$ and $\mu \leq \lambda$, then $\lambda = \mu$, and
- if $\lambda \leq \mu$ and $\mu \leq \nu$, then $\lambda \leq \nu$.

Question 5.3.1. How many partitions are $\leq (3, 3, 3, 3)$? Which ones are they?

Hasse diagrams

We can visualize partial orders by drawing their *Hasse diagrams*. A **Hasse diagram** is a graph $G = (V, E)$ where the vertices are the elements of the poset and there is an edge between distinct vertices a to b if and only if b covers a, i.e. $a \leq b$, and there are no distinct third elements c such that $a \leq c \leq b$. As graphs, Hasse diagrams are simple, directed, and acyclic.

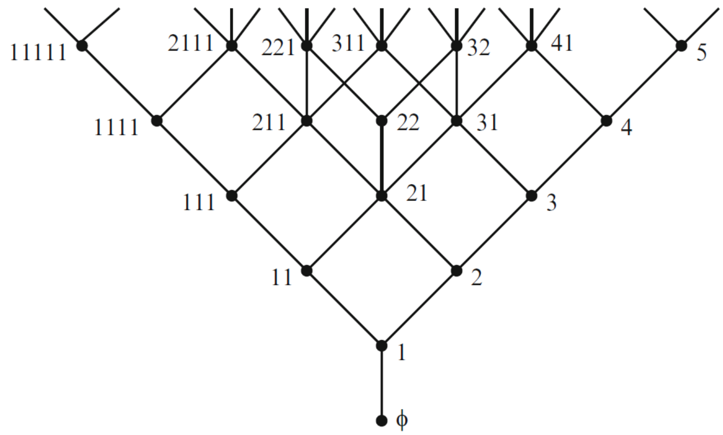
If we are interested in partial orders of partitions, then two partitions will be connected by an edge in the Hasse diagram if one can be obtained from the other by adding/removing a box. If we restrict our attention to only Young diagrams that fit in an $m \times n$ rectangle, we can draw finite Hasse diagrams corresponding to the partial order on these partitions. For example, depicted below are the Hasse diagrams for partitions that fit in 1×4, 2×2, and 2×3 rectangles, respectively.



Hasse diagrams are often drawn without directed edges, with the assumption that all edges are implicitly directed upwards.

Young's lattice

If we do not restrict the sizes of our partitions, we get an infinite partially-ordered set known as **Young's lattice**. The Hasse diagram is therefore an infinite graph, depicted below.



Let's look at walks starting at the empty partition in Young's lattice.[1] Since all edges are implicitly directed, we are starting at the bottom and only taking upward steps. A walk from the empty partition to λ can be thought of as building the partition λ by adding one block at a time.

If the Hasse diagram looked like a tree, then there would only be one way to reach each vertex from the empty partition, but it is not. For example, there are three distinct paths that we could take to reach $\lambda = (3, 1)$, representing the three distinct ways to make the Young diagram for $(3, 1)$ by adding one block at a time. If we number the blocks in each Young diagram based on the order that we added them, then we get three *Young tableaux* corresponding to $\lambda = (3, 1)$.

$$\begin{array}{|c|c|c|}\hline 1 & 2 & 3 \\ \hline 4 \\ \cline{1-1}\end{array} \qquad \begin{array}{|c|c|c|}\hline 1 & 3 & 4 \\ \hline 2 \\ \cline{1-1}\end{array} \qquad \begin{array}{|c|c|c|}\hline 1 & 2 & 4 \\ \hline 3 \\ \cline{1-1}\end{array}$$

In general, a **standard Young tableau** (SYT) is a Young diagram corresponding to a partition of n, where each box is labeled with a unique element of $[n]$ such that the numbers increase along rows and columns. The underlying partition λ corresponding to a Young tableau τ is called the **shape** of the tableau; this relationship is denoted $\lambda = \mathrm{sh}(\tau)$. As demonstrated above, there is a bijection between standard Young tableaux of shape λ and paths $\emptyset \to \lambda$ in Young's lattice.

Question 5.3.2. How many distinct standard Young tableaux of shape $(2, 2, 1)$ are there?

[1]This is the third page and we still have not defined Young tableaux, but don't give up; we are almost there!

Counting standard Young tableax

Given a partition $\lambda \vdash n$, how many distinct standard Young tableaux of shape λ are there? Considering that partitions are famously difficult to count, one may suspect that there is no nice answer to this question either. However, it turns out that we actually can count SYT of shape λ rather easily with the *hook-length formula*.

Let f^λ denote the number of SYT of shape λ. Given a box b in the Young diagram of λ, let $h(b)$ be the number of boxes that come after b in the same row or column, including b itself. This number $h(b)$ is called the **hook length** of b.

Theorem 5.3.3 (The hook-length formula). *If $\lambda \vdash n$, then $f^\lambda = \dfrac{n!}{\prod_{b\in\lambda} h(b)}$.*

For example, we have already calculated that the number of distinct SYT of shape $(2, 2, 1)$ is 5; let's use the hook-length formula to check our answer. We label each box of the Young diagram corresponding to $(2, 2, 1)$ by its hook length for convenience, then multiply all our numbers and divide 5! by the product.

$$\begin{array}{|c|c|}\hline (4) & (2) \\ \hline (3) & (1) \\ \hline (1) & \multicolumn{1}{c}{} \\ \cline{1-1}\end{array} \qquad f^{(2,2,1)} = \frac{5!}{4(2)(3)(1)(1)} = 5$$

Note that labeling each box by the hook length does *not* give us a SYT; parentheses are used here to indicate that this is not a valid labeling and is just useful for counting.

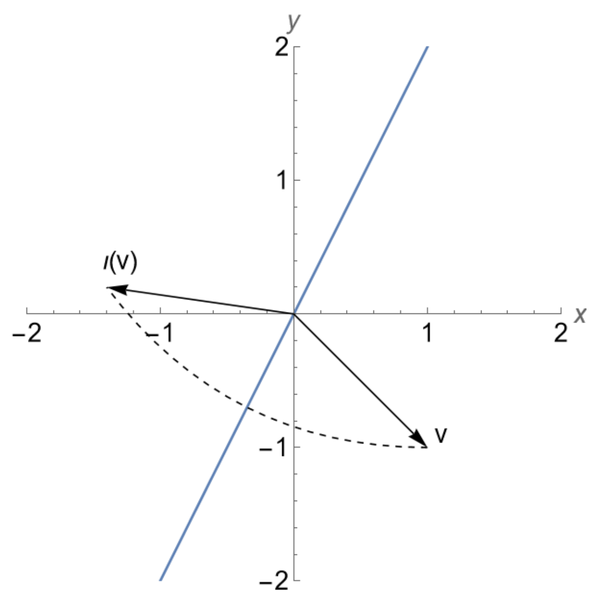
Question 5.3.4. Find a closed-form expression for the number of standard Young tableaux of shape (n, n).

5.4 Permutations and tableaux

Representation theory

Recall from linear algebra that a *(real) vector space* is a set V with two bits of structure: a vector addition map $+ : V \times V \to V$ and a scalar multiplication map $\cdot : \mathbb{R} \times V \to V$ satisfying certain properties. It turns out that every such finite-dimensional vector space is isomorphic to $\mathbb{R}^n$ for some $n \in \mathbb{N}$. Therefore, we can think of the one-dimensional vector space $\mathbb{R}^1$ as the single "building block" that can be combined with itself to make any space. Often, we are interested in vector spaces with more structure, such as a norm or an inner product. What are the building blocks of these more complicated structures?

For the sake of simplicity, let's talk about vector spaces V equipped with an **involution** $\iota : V \to V$, which is a linear transformation such that $\iota^2 = \mathrm{id}_V$, or equivalently $\iota(\iota(v)) = v$ for all $v \in V$. For example, consider the vector space $V = \mathbb{R}^2$ with the involution ι given by reflection across the line $y = 2x$.

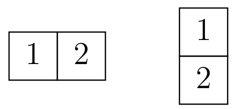


We can write ι in the standard basis $\mathcal{E}$ with matrix $\frac{1}{5}\begin{pmatrix} -3 & 4 \\ 4 & 3 \end{pmatrix}$. However, we can find a basis in which ι has a much simpler form! Using row/column operations, diagonalization, or just geometric reasoning, we find the basis $\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \end{pmatrix} \right\}$, in which ι is represented by the matrix $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. It turns out that any vector space V with an involution ι has a basis in which ι is represented by a diagonal matrix with only 1 and -1 as its nonzero entries. Therefore, we might say that there are *two* distinct building blocks of vector spaces with involutions: $\mathbb{R}^1$ with the identity involution $\begin{pmatrix} 1 \end{pmatrix}$, and $\mathbb{R}^1$ with the negation involution $\begin{pmatrix} -1 \end{pmatrix}$.

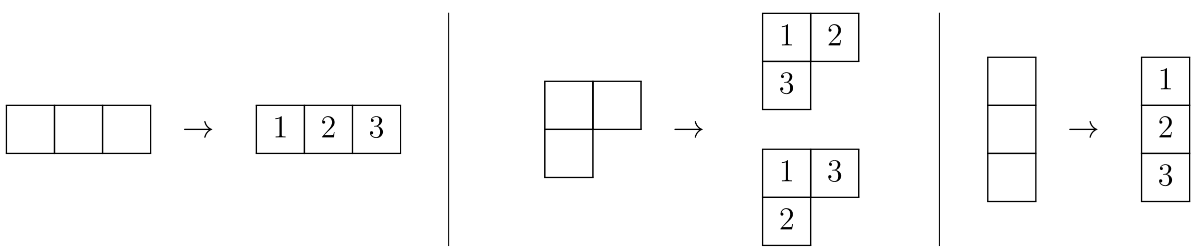
A "vector space with an involution" is a special case of a *representation of the symmetric group S_n*. Specifically, an involution is a representation of S_2. The technical term for a "building block" is an *irreducible representation*. Therefore, a restatement of the above is that there are two distinct irreducible representations of S_2: the *trivial representation* where ι is the identity, and the *standard representation* where ι is the negative of the identity.

Young tableaux in representation theory

How does this all relate to Young tableaux? The fact that there are two distinct building blocks of vector spaces with involutions corresponds to the fact that there are two distinct partitions of 2: (2) and $(1,1)$. Each building block is 1-dimensional, because there is only one distinct Young tableau of each shape:



A *representation of* S_3 is a vector space V with two extra pieces of structure: an involution $\iota : V \to V$ and another linear transformation $\rho : V \to V$ such that $\rho^3 = \mathrm{id}_V$ and $\iota\rho\iota = \rho^{-1}$. While we will not work this example out, we already have the ability to calculate how many building blocks a.k.a. irreducible representations of S_3 there are, and what dimension each one has. There are three distinct partitions of 3: (3), $(2,1)$, and $(1,1,1)$. The first and last of these partitions each corresponds to a unique Young tableaux, and the middle partition corresponds to two distinct Young tableaux. Therefore, S_3 has two irreducible 1-dimensional representations, and one irreducible 2-dimensional representation.



In general, representation theory as a whole, and the topic of symmetric groups in particular, has a lot of applications in other areas of math, physics, and computer science. While we won't go any deeper into this subject, it is worth noting that Young tableaux are an important part of this picture. The following theorem generalizes our examples above.

Theorem 5.4.1. *There is a bijection between irreducible representations of S_n and partitions of n. Furthermore, given $\lambda \vdash n$, the dimension of the representation corresponding to λ is f^λ, the number of Young tableaux of shape λ.*

The RSK correspondence

Fix some $n \in \mathbb{N}$: an interesting pattern emerges when we square the numbers f^λ for each $\lambda \vdash n$ and add them up. For example,

$$\begin{aligned}
&(f^{(1)})^2 &&= 1^2 &&= 1\\
&(f^{(2)})^2 + (f^{(1,1)})^2 &&= 1^2 + 1^2 &&= 2\\
&(f^{(3)})^2 + (f^{(2,1)})^2 + (f^{(1,1,1)})^2 &&= 1^2 + 2^2 + 1^2 &&= 6\\
&(f^{(4)})^2 + (f^{(3,1)})^2 + (f^{(2,2)})^2 + (f^{(2,1,1)})^2 + (f^{(1,1,1,1)})^2 &&= 1^2 + 3^2 + 2^2 + 3^2 + 1^2 &&= 24\,.
\end{aligned}$$

It turns out that this pattern holds for all n, i.e. it is true that

$$\sum_{\lambda \vdash n} (f^\lambda)^2 = n!\,.$$

How could we prove this? The **Robinson-Schensted-Knuth (RSK) correspondence** is a bijection from permutations of $[n]$ to pairs of tableaux with the same shape $\lambda \vdash n$. Starting with a permutation written in one line notation, like $\sigma = 4351276$, it reads the permutation left-to-right to construct two tableaux. While the output ends up being a standard Young tableaux, the intermediate steps involve **incomplete Young tableaux** (also called *near Young tableaux*). An incomplete tableau is one that still contains distinct entries that increase along rows and columns, but may contain only some of the numbers $1, \ldots, n$ instead of all of them.

Starting with a pair of empty tableaux $(\mathcal{P}, \mathcal{Q})$, the RSK algorithm takes a permutation $\sigma \in S_n$ and inserts the entries of σ into $\mathcal{P}$, called the **insertion tableau**, one at a time. Whenever the algorithm inserts σ_i into $\mathcal{P}$, it looks at the first row of $\mathcal{P}$ and finds the smallest entry greater than σ_i. If no such entry exists, then σ_i is placed at the end of the first row, and the insertion is complete. Otherwise, it replaces the greatest entry x smaller than σ_i with σ_i, and repeats this process by attempting to fit x into the second row. It continues fitting entries into lower rows and bumping the existing entries out until some entry is placed at the end of a row, and then the insertion is complete. Once σ_i is inserted into $\mathcal{P}$, a box labeled i is added to the **recording tableau** $\mathcal{Q}$ so that $\mathcal{Q}$ has the same shape as the updated $\mathcal{P}$. The first element inserted into the empty tableau is σ_1, then σ_2, and so on, until all entries of σ have been inserted into $\mathcal{P}$ and recorded on $\mathcal{Q}$.

Here is an example insertion of the number 6 into an incomplete tableau.

$$\begin{array}{|c|c|c|l}
\hline
1 & 2 & 7 & \leftarrow\ 6 \\
\hline
3 & 5 \\
\cline{1-2}
4 \\
\cline{1-1}
\end{array}
\qquad\Bigg|\qquad
\begin{array}{|c|c|c|l}
\hline
1 & 2 & 6 \\
\hline
3 & 5 & \multicolumn{1}{l}{\leftarrow\ 7} \\
\cline{1-2}
4 \\
\cline{1-1}
\end{array}
\qquad\Bigg|\qquad
\begin{array}{|c|c|c|}
\hline
1 & 2 & 6 \\
\hline
3 & 5 & 7 \\
\hline
4 \\
\cline{1-1}
\end{array}$$

For a full-size example, here are the steps taken by the RSK algorithm on input $\sigma = 4351276$.

$$\mathcal{P} = \emptyset \qquad \mathcal{Q} = \emptyset$$

$$\mathcal{P} = \begin{array}{|c|}\hline 4 \\ \hline\end{array} \qquad \mathcal{Q} = \begin{array}{|c|}\hline 1 \\ \hline\end{array}$$

$$\mathcal{P} = \begin{array}{|c|}\hline 3 \\ \hline 4 \\ \hline\end{array} \qquad \mathcal{Q} = \begin{array}{|c|}\hline 1 \\ \hline 2 \\ \hline\end{array}$$

$$\mathcal{P} = \begin{array}{|c|c|}\hline 3 & 5 \\ \hline 4 \\ \cline{1-1}\end{array} \qquad \mathcal{Q} = \begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 \\ \cline{1-1}\end{array}$$

$$\mathcal{P} = \begin{array}{|c|c|}\hline 1 & 5 \\ \hline 3 \\ \cline{1-1} 4 \\ \cline{1-1}\end{array} \qquad \mathcal{Q} = \begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 \\ \cline{1-1} 4 \\ \cline{1-1}\end{array}$$

$$\mathcal{P} = \begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 & 5 \\ \hline 4 \\ \cline{1-1}\end{array} \qquad \mathcal{Q} = \begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 & 5 \\ \hline 4 \\ \cline{1-1}\end{array}$$

$$\mathcal{P} = \begin{array}{|c|c|c|}\hline 1 & 2 & 7 \\ \hline 3 & 5 \\ \cline{1-2} 4 \\ \cline{1-1}\end{array} \qquad \mathcal{Q} = \begin{array}{|c|c|c|}\hline 1 & 3 & 6 \\ \hline 2 & 5 \\ \cline{1-2} 4 \\ \cline{1-1}\end{array}$$

$$\mathcal{P} = \begin{array}{|c|c|c|}\hline 1 & 2 & 6 \\ \hline 3 & 5 & 7 \\ \hline 4 \\ \cline{1-1}\end{array} \qquad \mathcal{Q} = \begin{array}{|c|c|c|}\hline 1 & 3 & 6 \\ \hline 2 & 5 & 7 \\ \hline 4 \\ \cline{1-1}\end{array}$$

This correspondence between permutations and pairs of SYT has some surprisingly nice properties!

- If σ corresponds to $(\mathcal{P}, \mathcal{Q})$, then σ^{-1} corresponds to $(\mathcal{Q}, \mathcal{P})$.
- If $\sigma = (\sigma_1, \ldots, \sigma_n)$ corresponds to $(\mathcal{P}, \mathcal{Q})$, then the reverse $(\sigma_n, \ldots, \sigma_1)$ corresponds to $(\mathcal{P}^*, \mathcal{Q}')$, where $\mathcal{P}^*$ is the conjugate Young tableau of $\mathcal{P}$, and $\mathcal{Q}'$ is also computable directly from $\mathcal{Q}$.
- The length of the longest increasing subsequence of σ is equal to the length of the first row of $\mathcal{P}$ (equivalently, $\mathcal{Q}$).
- Similarly, the length of the longest *decreasing* subsequence of σ is equal to the length of the first *column* of $\mathcal{P}$ (equivalently, $\mathcal{Q}$).

Bibliography

[Bog+17] K. Bogart et al. *Combinatorics Through Guided Discovery*. CreateSpace Independent Publishing Platform, 2017. ISBN: 9781981746590. URL: https://books.google.com/books?id=paChswEACAAJ.

[Bon16] M. Bona. *Walk Through Combinatorics, A: An Introduction To Enumeration And Graph Theory (Fourth Edition)*. World Scientific Publishing Company, 2016. ISBN: 9789813148864. URL: https://books.google.com/books?id=uZRIDQAAQBAJ.

[Mar01] G.E. Martin. *Counting: The Art of Enumerative Combinatorics*. Undergraduate Texts in Mathematics. Springer New York, 2001. ISBN: 9780387952253. URL: https://books.google.com/books?id=59xWF7hH%5C_9IC.

[SF14] R. Sedgewick and P. Flajolet. *An Introduction to the Analysis of Algorithms (2nd Edition)*. CreateSpace Independent Publishing Platform, 2014. ISBN: 9781502575869. URL: https://books.google.com/books?id=Re0woQEACAAJ.